

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM nucleic - protein search, using frame\_plus\_n2p model  
Run on: December 2, 2002, 20:34:50 ; Search time 23 Seconds  
(without alignments)  
2814.366 Million cell updates/sec

Title: US-09-896-791B-2  
Perfect score: 2010  
Sequence: 1 gaattcgacagggccat.....aaaaaaacatcgggcgcc 1100

Scoring table: BLOSUM62  
Xgapop 10.0, Xgapext 0.5  
Ygapop 10.0, Ygapext 0.5  
Fgapop 6.0, Fgapext 7.0  
Delop 6.0, Delext 7.0

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 525148

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame\_n2p.model -DEV=xlp  
-Q=/cgn2\_6/ptodata/2/iaa/5A.COMB.pdb:1287  
-DB=Issued Patents\_AA -QFMT=fastan -SUFFIX=ra1 -MINMATCH=0.1 -IOFCL=0  
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi  
-LIST=45 -DOCALIGN=200 -THR\_SCORE=pct -THR\_MAX=100 -THR\_MIN=0 -ALIGN=15  
-MODE=LOCAL -OUTFMT=pct -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US09896791.ecgn\_l1\_9 -runat\_02122002\_152659\_5635 -NCPU=6 -ICPU=3  
-NO\_XLPXY -NO\_MMAP -LARGEQUERY -NEG\_SCORES=0 -WAIT -LONGLOG -DEV\_TIMEOUT=120  
-WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7  
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Issued Patents\_AA: \*  
1: /cgn2\_6/ptodata/2/iaa/5A.COMB.pdb: \*  
2: /cgn2\_6/ptodata/2/iaa/5B.COMB.pdb: \*  
3: /cgn2\_6/ptodata/2/iaa/6A.COMB.pdb: \*  
4: /cgn2\_6/ptodata/2/iaa/6B.COMB.pdb: \*  
5: /cgn2\_6/ptodata/2/iaa/PCTUS.COMB.pdb: \*  
6: /cgn2\_6/ptodata/2/iaa/backfiles1.pdb: \*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	607.5	30.2	810	1	US-08-785-241-7 Sequence 7, Appl
2	606.5	30.2	245	4	US-09-438-833-3 Sequence 3, Appl
3	606.5	30.2	330	4	US-09-438-833-4 Sequence 4, Appl
4	606.5	30.2	652	4	US-09-438-833-5 Sequence 5, Appl
5	606.5	30.2	813	4	US-09-438-833-12 Sequence 12, Appl
6	606.5	30.2	826	1	US-08-785-241-6 Sequence 6, Appl
7	606.5	30.2	826	2	US-08-480-473B-2 Sequence 2, Appl
8	606.5	30.2	826	3	US-08-915-213-2 Sequence 2, Appl
9	606.5	30.2	826	3	US-09-148-547-2 Sequence 2, Appl
10	606.5	30.2	826	4	US-09-235-217-2 Sequence 2, Appl
11	606.5	30.2	826	4	US-09-380-662-23 Sequence 23, Appl
12	606.5	30.2	826	4	US-09-438-833-1 Sequence 1, Appl

13	606.5	30.2	826	5	PCT-US96-10251-2 Sequence 2, Appl
14	583	29.0	875	1	US-08-785-241-5 Sequence 5, Appl
15	571.5	28.4	485	4	US-09-374-454-4 Sequence 4, Appl
16	571.5	28.4	870	1	US-08-785-241-4 Sequence 4, Appl
17	571.5	28.4	870	4	US-09-374-454-6 Sequence 6, Appl
18	555.5	27.6	373	2	US-08-480-473B-3 Sequence 3, Appl
19	555.5	27.6	373	3	US-08-915-213-3 Sequence 3, Appl
20	555.5	27.6	373	4	US-09-235-217-3 Sequence 3, Appl
21	555.5	27.6	373	5	PCT-US96-10251-3 Sequence 3, Appl
22	555.5	27.6	805	2	US-08-480-473B-4 Sequence 4, Appl
23	555.5	27.6	805	3	US-08-915-213-4 Sequence 4, Appl
24	555.5	27.6	805	4	US-09-235-217-4 Sequence 4, Appl
25	555.5	27.6	805	5	PCT-US96-10251-4 Sequence 4, Appl
26	457	22.7	613	4	US-09-438-833-6 Sequence 6, Appl
27	412	20.5	756	4	US-09-438-833-11 Sequence 11, Appl
28	340.5	16.9	590	2	US-08-785-310A-5 Sequence 5, Appl
29	337	16.8	594	2	US-08-785-310A-6 Sequence 6, Appl
30	214	10.6	74	4	US-09-438-833-2 Sequence 2, Appl
31	203	10.1	54	2	US-08-480-473B-36 Sequence 36, Appl
32	203	10.1	54	3	US-08-915-213-36 Sequence 36, Appl
33	203	10.1	54	4	US-09-235-217-36 Sequence 36, Appl
34	199	9.9	54	4	US-09-374-454-3 Sequence 3, Appl
35	197	9.8	50	4	US-09-150-460B-17 Sequence 17, Appl
36	188	9.4	50	4	US-09-150-460B-18 Sequence 18, Appl
37	156.5	7.8	626	4	US-08-971-188-10 Sequence 10, Appl
38	156.5	7.8	789	4	US-08-971-188-9 Sequence 9, Appl
39	156.5	7.8	789	4	US-09-374-454-22 Sequence 22, Appl
40	154.5	7.7	626	4	US-09-374-454-21 Sequence 21, Appl
41	153	7.6	602	4	US-09-374-454-19 Sequence 19, Appl
42	145	7.2	808	4	US-08-971-188-12 Sequence 12, Appl
43	145	7.2	848	1	US-08-045-806-4 Sequence 4, Appl
44	145	7.2	848	1	US-08-366-051B-4 Sequence 4, Appl
45	142.5	7.1	51	4	US-09-150-460B-15 Sequence 15, Appl

ALIGNMENTS

RESULT 1  
US-08-785-241-7  
; Sequence 7, Application US/08785241  
; Patent No. 5695963  
; GENERAL INFORMATION:  
; APPLICANT: McKnight, Steven L.  
; APPLICANT: Russell, David W.  
; APPLICANT: Tian, Hui  
; TITLE OF INVENTION: Endothelial PAS Domain Protein  
; NUMBER OF SEQUENCES: 7  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: SCIENCE & TECHNOLOGY LAW GROUP  
; STREET: 268 BUSH STREET, SUITE 3200  
; CITY: SAN FRANCISCO  
; STATE: CALIFORNIA  
; COUNTRY: USA  
; ZIP: 94104  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/785,241  
; FILING DATE: 17-JAN-1997  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: OSMAN, RICHARD A  
; REGISTRATION NUMBER: 36,627  
; REFERENCE/DOCKET NUMBER: UTSD:1229  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (415) 343-4341  
; INFORMATION FOR SEQ ID NO: 7:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 810 amino acids

; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: peptide  
US-08-785-241-7

## Alignment Scores:

Pred. No.: 9,71e-52 Length: 810  
Score: 607.50 Matches: 128  
Percent Similarity: 67.65% Conservative: 33  
Best Local Similarity: 53.78% Mismatches: 49  
Query Match: 30.22% Indels: 28  
Diffs: 1 Caps: 3

US-09-896-791B-2 (1-1100) x US-08-785-241-7 (1-810)

QY 49 AACACCGAGCTGGGAGGAGAGTCCGGGAGCGCGCCGCGAGCGCGGCGAGGAG 108  
Db 2 SerSerGluArgArgLysGluLysSerArgAspAlaAlaArgSerArgArgThrLysGlu 21  
QY 109 ACGGAGTGTGTACCAAGCTGGCGCACACTCTGCCCTTTCCGCGCGCGCGCAC 168  
Db 22 SerGluValPheThrGluLeuAlaHisGlnLeuProLeuProHisAsnValSerHis 41  
QY 169 CTGGACAAAGCCTCATCATCGCCTCACAAATCAGCTTACCTGGCGCATCGACCGCTCTGC 228  
Db 42 LeuAspLysAlaSerValMetArgLeuThrIleSerThrLeuArgValArgLysLeu 61  
QY 229 GCAGCAGTGGAAAAGGGGAGAG-----CCACTGGACGCCCTGTACCT 273  
Db 62 AspAlaGlyGly-LeuAspSerGluAspGluMetLysAlaGlnMetAspCysPheThrLe 81  
QY 274 GAAGGCCCTGGAGGTTTCGTCTCATGTGTACTCACCGCGGAGGAGACATGCTTACCTGTC 333  
Db 81 uLysAlaLeuAspGlyPheValMetValLeuThrAspAspGlyAspMetValThrIleSe 101  
QY 334 GGAATATGTTCAGCAAGACCTGGCGCCTCAGTCACTGGAGACCTCTGTCTCTCCCTCCCTGATA 393  
Db 101 rAspAsnValAsnLysThrMetGlyLeuThrGln----- 112  
QY 394 CATACCCCACTCTGTGTACCAATTTCTCTGTGAGCTCATTTGGACACATATCTTTGAT 453  
Db 113 -----Phe--GluLeuAlaGlyHisSerValPheAsp 122  
QY 454 TTTATCCATCCCTGTGACCAAGAGAACTTCAAGACGCCCTGACCCCGAGCGCGAACCTG 513  
Db 123 PheThrHisProCysAspHisGluGluMetArgGluMetLeuThrHisArgAsnGlyPro 142  
QY 514 TCAAGAAGAAGCTGGAAGCCCCAACAGAGCGCCACTTTTCCCTGCGAATGAAGAGCAGC 573  
Db 143 ValArgLysGlyLysGluLeuAsnThrGlnArgSerPhePheLeuArgMetLysCysThr 162  
QY 574 CTACACGACAGAGGGCGCGCTCAACCTCAAGCGCGCACCTGGAGAGGTGCTGCACTGC 633  
Db 163 LeuThrSerArgGlyArgThrMetAsnIleLysSerAlaThrTrpLysValLeuHisCys 182  
QY 634 TCAGGACATATGAGGGCTTACAAGCCCCCTGCACAGACTTCCCTCGCGCGCGCCCTCGC 693  
Db 183 ThrGlyHisIleHisValThrAspThrAsnSerAsnGlnProGlnCysGly-----Tyr 200  
QY 694 TCCGAGCCTCCCTGCAATCCCTGGTGTCTTATCTGTGAAGCCATCCCC 741  
Db 201 LysLysProProMetThrCysLeuValLeuIleCysGluProIlePro 216

## RESULT 2

US-09-438-833-3  
; Sequence 3, Application US/09438833  
; Patent No. 6436654  
; GENERAL INFORMATION:  
; APPLICANT: Pharmacia & Upjohn  
; TITLE OF INVENTION: Protein variants  
; FILE REFERENCE: 1848  
; CURRENT APPLICATION NUMBER: US/09/438,833

; CURRENT FILING DATE: 1999-11-12  
; NUMBER OF SEQ ID NOS: 15  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 3  
; LENGTH: 245  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: Subdomain  
; OTHER INFORMATION: 1-245 of human HIF-1 alpha  
US-09-438-833-3

## Alignment Scores:

Pred. No.: 7,56e-52 Length: 245  
Score: 606.50 Matches: 129  
Percent Similarity: 68.07% Conservative: 33  
Best Local Similarity: 54.20% Mismatches: 48  
Query Match: 30.17% Indels: 28  
Diffs: 4 Caps: 3

US-09-896-791B-2 (1-1100) x US-09-438-833-3 (1-245)

QY 49 AACACCGAGCTGGGAGGAGAGTCCGGGAGCGCGCCGCGAGCGCGGCGAGGAG 108  
Db 14 SerSerGluArgArgLysGluLysSerArgAspAlaAlaArgSerArgArgSerLysGlu 33  
QY 109 ACGGAGTGTGTACCAAGCTGGCGCACACTCTGCCCTTTCCGCGCGCGCGCAC 168  
Db 34 SerGluValPheThrGluLeuAlaHisGlnLeuProLeuProHisAsnValSerHis 53  
QY 169 CTGGACAAAGCCTCATCATCGCCTCACAAATCAGCTTACCTGGCGCATCGACCGCTCTGC 228  
Db 54 LeuAspLysAlaSerValMetArgLeuThrIleSerThrLeuArgValArgLysLeu 73  
QY 229 GCAGCAGTGGAAAAGGGGAGAG-----CCACTGGACGCCCTGTACCT 273  
Db 74 AspAlaGly-LeuAspSerGluAspMetLysAlaGlnMetAsnCysPheThrLe 93  
QY 274 GAAGGCCCTGGAGGTTTCGTCTCATGTGTACTCACCGCGGAGGAGACATGCTTACCTGTC 333  
Db 93 uLysAlaLeuAspGlyPheValMetValLeuThrAspAspGlyAspMetIleThrIleSe 113  
QY 334 GGAATATGTTCAGCAAGACCTGGCGCCTCAGTCACTGGAGACCTCTGTCTCTCCCTCCCTGATA 393  
Db 113 rAspAsnValAsnLysThrMetGlyLeuThrGln----- 124  
QY 394 CATACCCCACTCTGTGTACCAATTTCTCTGTGAGCTCATTTGGACACATATCTTTGAT 453  
Db 125 -----Phe--GluLeuThrGlyHisSerValPheAsp 134  
QY 454 TTTATCCATCCCTGTGACCAAGAGAACTTCAAGACGCCCTGACCCCGAGCGCGAACCTG 513  
Db 135 PheThrHisProCysAspHisGluGluMetArgGluMetLeuThrHisArgAsnGlyLeu 154  
QY 514 TCAAGAAGAAGCTGGAAGCCCCAACAGAGCGCCACTTTTCCCTGCGAATGAAGAGCAGC 573  
Db 155 ValLysLysGlyLysGluGlnAsnThrGlnArgSerPhePheLeuArgMetLysCysThr 174  
QY 574 CTACACGACAGAGGGCGCGCTCAACCTCAAGCGCGCACCTGGAGAGGTGCTGCACTGC 633  
Db 175 LeuThrSerArgGlyArgThrMetAsnIleLysSerAlaThrTrpLysValLeuHisCys 194  
QY 634 TCAGGACATATGAGGGCTTACAAGCCCCCTGCACAGACTTCCCTCGCGCGGAGCCCTCGC 693  
Db 195 ThrGlyHisIleHisValThrAspThrAsnSerAsnGlnProGlnCysGly-----Tyr 212  
QY 694 TCCGAGCCTCCCTGCAATCCCTGGTGTCTTATCTGTGAAGCCATCCCC 741  
Db 213 LysLysProProMetThrCysLeuValLeuIleCysGluProIlePro 228

## RESULT 3

US-09-438-833-4  
; Sequence 4, Application US/09438833

```

; Patent No. 6436654
; GENERAL INFORMATION:
; APPLICANT: Pharmacia & Upjohn
; TITLE OF INVENTION: Protein variants
; FILE REFERENCE: 1848
; CURRENT APPLICATION NUMBER: US/09/438,833
; CURRENT FILING DATE: 1999-11-12
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4
; LENGTH: 330
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Subdomain
; OTHER INFORMATION: 1-330 Of human HIF-1 alpha
US-09-438-833-4

Alignment Scores:
Pred. No.:      8,51e-52      Length:      330
Score:          606.50      Matches:    129
Percent Similarity: 68.07%      Conservative: 33
Best Local Similarity: 54.20%      Mismatches: 48
Query Match:    30.17%      Indels:     28
DB:             4          Gaps:         3

US-09-896-791B-2 (1-1100) x US-09-438-833-4 (1-330)
QY 49 ACACCGAGCTCGGAAGAGAGAGTCCGGGAGCGGGCCGCGGAGCGGCGGAGCGAGGAG 108
Db 14 SerSerGluArgArgLysGluLysSerArgAspAlaAlaArgSerArgArgSerLysGlu 33
QY 109 ACGGAGGTCTCTACAGCTGCGGCACACTCTGCCCTTTGCGCGGCGCTCAGCGGCAC 168
Db 34 SerGluValPheTyrGluLeuAlaHisGlnLeuProLeuProHisAsnValSerSerHis 53
QY 169 CTGGCAAGGCTCCATCATGCGCTCACAATCAGTCTCGGATCGGATCGGATCGCTCTGC 228
Db 54 LeuAspLysAlaSerValMetArgLeuThrIleSerTyrLeuArgValArgLysLeuLeu 73
QY 229 GCAGCAGGTGGAAAGGGGGAGAG-----CCACTGGAGCGCTCTGTCTCTCCCTCGTACT 273
Db 74 AspAlaGly-AspLeuAspIleGluAspMetLysAlaGlnMetAsnCysPheTyrLe 93
QY 274 GAAGGCCCTGGAGGTTTCGTCTACTGTCTACCGCGGAGGAGACATCGCTTACCTGTC 333
Db 93 uLysAlaLeuAspGlyPheValMetValLeuThrAspAspGlyAspMetIleTyrIleSe 113
QY 334 GGAATATGTCAAGACACCTGGGCTCAGTCAGTGGACCTCTGTCTCTCTCCCTCGTATA 393
Db 113 rAspAsnValAsnLysTyrMetGlyLeuThrGln----- 124
QY 394 CATACCCCATCTCTGGTACCAATTTCTCTCTGGAGCTCATTTGGACACAGTATCTTTGAT 453
Db 125 -----Phe--GluLeuThrGlyHisSerValPheAsp 134
QY 454 TTTATCCATCCTGTGACCAAGAGAACTTCAAGAGCGCTCGACCCCGGAGCGGACCTGTC 513
Db 135 PheThrHisProCysAspHisGluMetArgGluMetLeuThrHisArgAsnGlyLeu 154
QY 514 TCAAGAAGAAGCTGGAAGCCCCCAACAGAGCGCCACTTTTCCCTCGGAATGAAGACGAG 573
Db 155 ValLysLysGlyLysGluGlnAsnThrGlnArgSerPhePheLeuArgMetLysCysThr 174
QY 574 CTCACGAGGAGGCGGCGACGCTCAACCTCAAGCGGCCACCTGGAAGTGTGCTGCACTGC 633
Db 175 LeuThrSerArgGlyArgThrMetAsnIleLysSerAlaThrTrpLysValLeuHisCys 194
QY 634 TCAGGACATATAGGCGCTTACAAGCCCCCTCGCACAGACTTCCCTCCCGGAGCCCTCCG 693
Db 195 ThrGlyHisIleHisValTyrAspThrAsnSerAsnGlnProGlnCysGly-----Tyr 212
QY 694 TCCGAGCCTCCCTGCAATGCGTGTGCTTATCTGTGAAGCCATCCCC 741

```

```

Db 213 LysLysProPrometThrCysLeuValLeuIleCysGluProIlePro 228
RESULT 4
US-09-438-833-5
; Sequence 5, Application US/09438833
; Patent No. 6436654
; GENERAL INFORMATION:
; APPLICANT: Pharmacia & Upjohn
; TITLE OF INVENTION: Protein variants
; FILE REFERENCE: 1848
; CURRENT APPLICATION NUMBER: US/09/438,833
; CURRENT FILING DATE: 1999-11-12
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 5
; LENGTH: 652
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Subdomain
; OTHER INFORMATION: 1-652 Of human HIF-1 alpha
US-09-438-833-5

Alignment Scores:
Pred. No.:      1.12e-51      Length:      652
Score:          606.50      Matches:    129
Percent Similarity: 68.07%      Conservative: 33
Best Local Similarity: 54.20%      Mismatches: 48
Query Match:    30.17%      Indels:     28
DB:             4          Gaps:         3

US-09-896-791B-2 (1-1100) x US-09-438-833-5 (1-652)
QY 49 ACACCGAGCTCGGAAGAGAGAGTCCGGGAGCGGGCCGCGGAGCGGCGGAGCGGAG 108
Db 14 SerSerGluArgArgLysGluLysSerArgAspAlaAlaArgSerArgArgSerLysGlu 33
QY 109 ACGGAGGTCTCTACAGCTGCGGCACACTCTGCCCTTTGCGCGGCGCTCAGCGGCAC 168
Db 34 SerGluValPheTyrGluLeuAlaHisGlnLeuProLeuProHisAsnValSerSerHis 53
QY 169 CTGGCAAGGCTCCATCATGCGCTCACAATCAGTCTCGGATCGGATCGGATCGCTCTGC 228
Db 54 LeuAspLysAlaSerValMetArgLeuThrIleSerTyrLeuArgValArgLysLeuLeu 73
QY 229 GCAGCAGGTGGAAAGGGGGAGAG-----CCACTGGAGCGCTCTGTCTCTCCCTCGTACT 273
Db 74 AspAlaGly-AspLeuAspIleGluAspMetLysAlaGlnMetAsnCysPheTyrLe 93
QY 274 GAAGGCCCTGGAGGTTTCGTCTACTGTCTACCGCGGAGGAGACATCGCTTACCTGTC 333
Db 93 uLysAlaLeuAspGlyPheValMetValLeuThrAspAspGlyAspMetIleTyrIleSe 113
QY 334 GGAATATGTCAAGACACCTGGGCTCAGTCAGTGGACCTCTGTCTCTCTCCCTCGTATA 393
Db 113 rAspAsnValAsnLysTyrMetGlyLeuThrGln----- 124
QY 394 CATACCCCATCTCTGGTACCAATTTCTCTCTGGAGCTCATTTGGACACAGTATCTTTGAT 453
Db 125 -----Phe--GluLeuThrGlyHisSerValPheAsp 134
QY 454 TTTATCCATCCTGTGACCAAGAGAACTTCAAGAGCGCTCGACCCCGGAGCGGACCTGTC 513
Db 135 PheThrHisProCysAspHisGluMetArgGluMetLeuThrHisArgAsnGlyLeu 154
QY 514 TCAAGAAGAAGCTGGAAGCCCCCAACAGAGCGCCACTTTTCCCTCGGAATGAAGACGAG 573
Db 155 ValLysLysGlyLysGluGlnAsnThrGlnArgSerPhePheLeuArgMetLysCysThr 174
QY 574 CTCACGAGGAGGCGGCGACGCTCAACCTCAAGCGGCCACCTGGAAGTGTGCTGCACTGC 633
Db 175 LeuThrSerArgGlyArgThrMetAsnIleLysSerAlaThrTrpLysValLeuHisCys 194

```

```
QY 634 TCAGGACATATGAGGCGCTACAAGCCCTGCACAGACTTCCCTGCGGGAGGCCCTCGC 693
    ::::::::::: |||
Db 195 ThrGlyHisIleHisValTyrAspThrAsnSerAsnGlnProGlnCysGly-----Tyr 212
    ::::::::::: |||
QY 694 TCCGAGCTCCCTGCATGCTGGTCTTATCTGTGAAGCCATCCCC 741
    ::::::::::: |||
Db 213 LysLysProMetThrCysLeuValLeuIleCysGluProIlePro 228
    ::::::::::: |||
RESULT 5
US-09-438-833-12
; Sequence 12, Application US/09438833
; Patent No. 6436654
; GENERAL INFORMATION:
; APPLICANT: Pharmacia & Upjohn
; TITLE OF INVENTION: Protein variants
; FILE REFERENCE: 1848
; CURRENT APPLICATION NUMBER: US/09/438,833
; CURRENT FILING DATE: 1999-11-12
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 12
; LENGTH: 813
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Subdomain
; OTHER INFORMATION: 1-813 of human HIF-1 alpha
US-09-438-833-12

Alignment Scores:
Pred. No.: 1,22e-51 Length: 813
Score: 606.50 Matches: 129
Percent Similarity: 68.07% Conservative: 33
Best Local Similarity: 54.20% Mismatches: 48
Query Match: 30.17% Indels: 28
DB: 4 Gaps: 3

US-09-896-791B-2 (1-1100) x US-09-438-833-12 (1-813)
QY 49 AACACCGAGTCGGAGGAGAGTCGCGGAGCGCGCGCGGCGGCGGCGGCGGCGGAG 108
    ::::::::::: |||
Db 14 SerSerGluArgArgLysGluLysSerArgAspAlaAlaArgSerArgSerLysGlu 33
    ::::::::::: |||
QY 109 ACGGAGTGTGTACAGCTGCGCGCACACCTCTGCCCTTTTCGCGCGCGGCGGCGGCGCAC 168
    ::::::::::: |||
Db 34 SerGluValPheTyrGluLeuAlaHisGlnLeuProLeuProHisValSerSerHis 53
    ::::::::::: |||
QY 169 CTGACAAAGCCCTCCATCATGCGCCTCACAAATCAGCTACCTGCGCATGCGCCCTCTGC 228
    ::::::::::: |||
Db 54 LeuAspLysAlaSerValMetArgLeuThrIleSerTyrLeuArgValArgLysLeuLeu 73
    ::::::::::: |||
QY 229 GCACGAGTGGAAAAGGGGGAGAG-----CCACTGGAGCCCTGCTACCT 273
    ::::::::::: |||
Db 74 AspAlaGly-AspLeuAspIleGluAspMetLysAlaGlnMetAsnCysPheTyrLe 93
    ::::::::::: |||
QY 274 GAAGGCGCTGGAGGTTTCGTCTGTTGATCATACCGCGGAGGAGACATGCTTACCTGTC 333
    ::::::::::: |||
Db 93 uLysAlaLeuAspGlyPheValMetValLeuThrAspAspGlyAspMetIleTyrIleSe 113
    ::::::::::: |||
QY 334 GGAATATGTCAAGAACACCTGGCGCTCAGTCACTGAGACCTCTGTTCCCTCCCTCGATA 393
    ::::::::::: |||
Db 113 rAspAsnValAsnLysTyrMetGlyLeuThrGln----- 124
    ::::::::::: |||
QY 394 CATAACCCCACTCCTGGTACCATTCTCTCTGAGCTCATTTGGACACACATATCTTTGAT 453
    ::::::::::: |||
Db 125 -----Phe--GluLeuThrGlyHisSerValPheAsp 134
    ::::::::::: |||
QY 454 TTTATCCATCCCTGTGACCAAGAGAACTTCAAGACGCCCTGACCCCGGCGGCGGACCTG 513
    ::::::::::: |||
Db 135 PheThrHisProCysAspHisGluGluMetArgGluMetLeuThrHisArgAsnGlyLeu 154
    ::::::::::: |||
QY 514 TCAAAGAAGAAGCTGGAAGGCCCCAACAGAGCGCCACTTTTCCCTGCGGAATGAAGACAG 573
    ::::::::::: |||
```

```
Db 155 ValLysLysGlyLysGluGlnAsnThrGlnArgSerPhePheLeuArgMetLysCysThr 174
    ::::::::::: |||
QY 574 CTCACGAGCAGAGCGGCGGCGCTCAACCTCAAGCGGCGGCGGCGGCGGCGGCGGCTGC 633
    ::::::::::: |||
Db 175 LeuThrSerArgGlyArgThrMetAsnIleLysSerAlaThrTrpLysValLeuHisCys 194
    ::::::::::: |||
QY 634 TCAGGACATATGAGGCGCTACAAGCCCTGCACAGACTTCCCTGCGGGAGGCCCTCGC 693
    ::::::::::: |||
Db 195 ThrGlyHisIleHisValTyrAspThrAsnSerAsnGlnProGlnCysGly-----Tyr 212
    ::::::::::: |||
QY 694 TCCGAGCTCCCTGCATGCTGGTCTTATCTGTGAAGCCATCCCC 741
    ::::::::::: |||
Db 213 LysLysProMetThrCysLeuValLeuIleCysGluProIlePro 228
    ::::::::::: |||
RESULT 6
US-08-785-241-6
; Sequence 6, Application US/08785241
; Patent No. 5695963
; GENERAL INFORMATION:
; APPLICANT: McKnight, Steven L.
; APPLICANT: Russell, David W.
; APPLICANT: Tian, Hui
; TITLE OF INVENTION: Endothelial PAS Domain Protein
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SCIENCE & TECHNOLOGY LAW GROUP
; STREET: 268 BUSH STREET, SUITE 3200
; CITY: SAN FRANCISCO
; STATE: CALIFORNIA
; COUNTRY: USA
; ZIP: 94104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/785,241
; FILING DATE: 17-JAN-1997
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: OSMAN, RICHARD A
; REGISTRATION NUMBER: 36,627
; REFERENCE/DOCKET NUMBER: UTSD:1229
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 343-4341
; TELEFAX: (415) 343-4342
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 826 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-08-785-241-6

Alignment Scores:
Pred. No.: 1,23e-51 Length: 826
Score: 606.50 Matches: 129
Percent Similarity: 68.07% Conservative: 33
Best Local Similarity: 54.20% Mismatches: 48
Query Match: 30.17% Indels: 28
DB: 1 Gaps: 3

US-09-896-791B-2 (1-1100) x US-08-785-241-6 (1-826)
QY 49 AACACCGAGTCGGAGGAGAGTCGCGGAGCGCGCGGCGGCGGCGGCGGCGGAG 108
    ::::::::::: |||
Db 14 SerSerGluArgArgLysGluLysSerArgAspAlaAlaArgSerArgSerLysGlu 33
    ::::::::::: |||
QY 109 ACGGAGTGTGTACAGCTGCGCGCACACCTCTGCCCTTTTCGCGCGCGGCGGCGGCGCAC 168
    ::::::::::: |||
```

```
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 826 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-480-473B-2

Alignment Scores:
Pred. No.:      1,23e-51
Score:          606.50
Percent Similarity: 68.07%
Best Local Similarity: 54.20%
Query Match:     30.17%
DB:              2
Gaps:            3

US-09-896-791B-2 (1-1100) x US-08-480-473B-2 (1-826)

QY   49 AACACCGAGCTGGCGAAGGAGAAGTCGCGGAGCGCGCCGCAGCGCGCAGCCAGGAG 108
    ::::||||| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db   14 SerSerGluArgArgLysGluLysSerArgAspAlaAlaArgSerArgArgSerLysGlu 33
QY   109 ACGAGGTGCTGTACCAGCTGGCGCACACTCTGCCCTTTCCGCGCGCGCTCAGCGCCAC 168
    ::||| || | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db   34 SerGluValPheTyrGluLeuAlaHisGlnLeuProLeuHisAsnValSerSerHis 53
QY   169 CTGCACAAGCGCTCCATCATCGCGCTCACAACTACGCTACCTGCGCATGCACCGCCTCTGC 238
    ||||| ||| | | | | | | | | | | | | | | | | | | | | | | | | | |
Db   54 LeuAspLysAlaSerValMetArgLeuThrIleSerTyrLeuArgValArgLysLeuLeu 73
QY   229 GCACGAGGTGGAAAAGGGGGAGAG-----CCACTGGACCGCTCTACCT 273
    ||||| ::| ||| ||| :::: |||||
Db   74 AspAlaGly-AspLeuAspIleGluAspAspMetLysAlaGlnMetAsnCysPheTyrLe 93
QY   274 GAAGGCCCTGGAGGTTTCGTATGTTACTACCGCGCGAGGAGACATGCTTACCTGTCTC 333
    ||||| ||| | | | | | | | | | | | | | | | | | | | | | | | | | |
Db   93 uLysAlaLeuAspGlyPheValMetValLeuThrAspAspGlyAspMetIleTyrIleSe 113
QY   334 GGAATAATGTCAAGAACACCTGGGCGCTCAGTCAGTGAGACCTCTGTTCTCTCCCTGATA 393
    ||::| ||||| | | | | | | | | | | | | | | | | | | | | | | | |
Db   113 rASpAsnValAsnLysTyrMetGlyLeuThrGln----- 124
QY   394 CATACACCCCACTCCTGTGTACCAATTCTCTCTGAGCTCATTTGGACACAGTATCTTTGAT 453
    ||||| ||||| ||||| ||||| | | | | | | | | | | | | | | | | |
Db   125 -----Phe--GluLeuThrGlyHisSerValPheAsp 134
QY   454 TTTATCATCTCCTGTGACCAAGAGGAACCTCAAGACGCCCTGACCCCCAGCGCAACCTG 513
    || | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db   135 PheThrHisProCysAspHisGluLeuMetArgGluMetLeuThrHisArgAsnGlyLeu 154
QY   514 TCAAAGAAGAGCTGGAGCGCCCAACAGACGCCACTTTTCCCTGCCAANTGAAGACGACG 573
    ||||| ||||| | | | | | | | | | | | | | | | | | | | | | | |
Db   155 ValLysLysGlyLysGluGlnAsnThrGlnArgSerPhePheLeuArgMetLysCysThr 174
QY   574 CTCACCACGAGCGGCGCAGCTCAACCTCAAAGCGGCCACCTCGAAGGTGCTGCATCTGC 633
    ||||| ||||| | | | | | | | | | | | | | | | | | | | | | | |
Db   175 LeuThrSerArgGlyArgThrMetAsnIleLysSerAlaThrTrpLysValLeuHisCys 194
QY   634 TCAGACATATGAGGGGCTTCAAGCCCCCTGCACAGACTTCCCTCTGCCGGGAGCCCTCCG 693
    ::||| ||||| | | | | | | | | | | | | | | | | | | | | | |
Db   195 ThrGlyHisIleHisValTyrAspThrAsnSerAsnGlnProGlnCysGly-----Tyr 212
QY   694 TCCGAGCTCCCTGCGAATGCCCTGGTCTTATCTGTGAAGCCATCCCC 741
    ::||| ||||| | | | | | | | | | | | | | | | | | | | | | |
Db   213 LysLysProMetThrCysLeuValIleCysGluProIlePro 228

RESULT 8
US-08-915-213-2
; Sequence 2, Application US/08915213
; Patent No. 6020462
; GENERAL INFORMATION:
; APPLICANT: Semenza, Gregg L.
; TITLE OF INVENTION: HYPOXIA INDUCIBLE FACTOR-1 AND METHOD OF USE
```

; NUMBER OF SEQUENCES: 64  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Fish & Richardson P.C.  
 ; STREET: 4225 Executive Square, Suite 1400  
 ; CITY: La Jolla  
 ; STATE: CA  
 ; COUNTRY: USA  
 ; ZIP: 92037  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: PatentIn Release #1.0, Version #1.30  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/915,213  
 ; FILING DATE: 20-AUG-1997  
 ; CLASSIFICATION: 514  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: US 08/480,473  
 ; FILING DATE: 06-JUN-1995  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Haile, Lisa A.  
 ; REGISTRATION NUMBER: 38,347  
 ; REFERENCE/DOCKET NUMBER: 07265/053001  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: 619/678-5070  
 ; TELEFAX: 619/678-5099  
 ; INFORMATION FOR SEQ ID NO: 2:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 826 amino acids  
 ; TYPE: amino acid  
 ; TOPOLOGY: linear  
 ; MOLECULE TYPE: protein  
 ; US-08-915-213-2

Alignment Scores:	1.23e-51	Length:	826
Pred. No.:	606.50	Matches:	129
Score:	68.07%	Conservative:	33
Percent Similarity:	54.20%	Mismatches:	48
Best Local Similarity:	30.17%	Indels:	28
Query Match:	3	Gaps:	3
DB:			

US-09-896-791B-2 (1-1100) x US-08-915-213-2 (1-826)

QY	49	AACACCGAGTGC	CGGAAGAGAA	AGTCGCGGAG	CGCGCCCGCAGCGCGCGCAGCGAGG	108
Db	14	Ser-Glu	ArgArgIys	GluLysSerArg	AspAlaArgSerArgArgSerIysGlu	33
QY	109	ACGGAGGTGCTG	TACACAGCTGG	CGGCACACTG	CCCTTTGCGCGCGGCGTCACGCGGCAC	168
Db	34	Ser-Glu	ValPheTyr	GluLeuAlaHis	GlnLeuProLeuProHisAsnValSerSerHis	53
QY	169	CTGGACAAGGCC	TCCATCATGCGCC	CTCAATCAGCTAC	TGCGCATGACCCGCTCTGCTGC	228
Db	54	LeuAsp	LysAlaSerVal	MetArgLeuThrIle	SerTyrLeuArgValArgIysLeuLeu	73
QY	229	GCAGCAGGTG	GGAAGGGGGAG	-----	CCACTGGAGCGCTGCTACCT	273
Db	74	AspAla	agIy-AspLeu	AspIleGluAspMet	LysAlaGlnMetAsnCysPheTyrIle	93
QY	274	GAAGGCCCTG	AGGGTTTCGT	CATGTACTACCG	CCGAGGAGACATGGCTTACCTGTC	333
Db	93	uLysAlaLeu	AspGlyPheVal	MetValLeuThr	AspAspGlyAspMetIleTyrIleSe	113
QY	334	GGAATATGT	CAGCAAGCAC	CTGGCCCTCAG	TCCAGTGGACCTCTGTCTCCTCCCTGATA	393
Db	113	rasAsp	AsnValAsnLys	TyrMetGlyLeu	ThrGln-----	124
QY	394	CATAACCC	CACTCCTGGT	ACCAATTTCT	CTCGGAGCTCATTTGGACACAGTATCTTTGAT	453
Db	125	-----	-----	-----	Phe--GluLeuThrGlyHisSerValPheAsp	134

Qy	454	TTTATTCATCCCTGTGACCAAGAGAACTTCAAGACGCCCTTGACCCCCCGGCGGAACCTG	51
Db	135	PheThrHisProCysAspHisGluMetArgGluMetLeuThrHisArgAsnGlyLeu	111
Qy	514	TCAAAGAAGAACGTGGGAAGCCCACAACAGACGCCACTTTTCCCTGCGAATGAAGACGACG	573
Db	155	VallLysLysGlyIysGluInAsnThrGlnArgSerPhePheLeuArgMetLysCysThr	174
Qy	574	CTCACACGACGAGGGCGCACGCCTCAACCTCAAAGCGCCACCCTGGGAAGGTGTGCACTGC	633
Db	175	LeuThrSerArgGlyArgThrMetAsnIleLysSerAlaThrTrpLysValLeuHisCys	194
Qy	634	TACGACATATGAGGGGCTATAAGCCCCCTGCACAGACTTCCCTCCCGGGAGCCCTCCG	693
Db	195	ThrglyHisIleHisValTyAspThrAsnSerAsnGlnProGlnCysGly-----Tyr	212
Qy	694	TCCGAGCCTCCCTGCACCATCGCTGGTGTCTTATCTGTGAAGCCATCCCC	741
Db	213	LysLysProMetThrCysLeuValLeulleCysGluProilePro	228
<b>RESULT 9</b>			
US-09-148-547-2			
; Sequence 2, Application US/09148547			
; Patent No. 6124131			
; GENERAL INFORMATION:			
; APPLICANT: Semenza, Gregg L.			
; TITLE OF INVENTION: Hypoxia Inducible Factor-1 and Methods of Use			
; FILE REFERENCE: 07265/151001			
; CURRENT APPLICATION NUMBER: US/09/148,547			
; CURRENT FILING DATE: 1998-08-25			
; NUMBER OF SEQ ID NOS: 2			
; SOFTWARE: PatentIn Ver. 2.0			
; SEQ ID NO 2			
; LENGTH: 826			
; TYPE: PRT			
; ORGANISM: Homo sapiens			
US-09-148-547-2			
Alignment Scores:			
Pred. No.: 1,23e-51 Length: 826			
Score: 606.50 Matches: 129			
Percent Similarity: 68.07% Conservative: 33			
Best Local Similarity: 54.20% Mismatches: 48			
Query Match: 30.17% Indels: 28			
DB: Gaps: 3			
US-09-896-791B-2 (1-1100) x US-09-148-547-2 (1-826)			
Qy	49	AACACCGAGCTGCGGAAGAGAAGTCGCGGACGCGCGCCGCGAGCGCGCGACCGAGG	108
Db	14	SerSerGluArgArgLysGluLysSerArgAspAlaArgSerArgArgSerLysGlu	33
Qy	109	ACGAGGTGCTGTACACAGTGGCGCACACTCTCCCTTTGGCGGCGCTCAGCGCGAC	168
Db	34	SerGluValPheTyrgluLeuAlaHisGlnLeuProLeuProHisasnValSerSerHis	53
Qy	169	CTGCACAAGCCCTCCATCATGCGCCCTCACAACTACGTACTCTCGCATGCACCGCTCTGC	228
Db	54	LeuAspLysAlaSerValMetArgLeuthrIleSerTyrgluArgLysValArgLysLeu	73
Qy	229	GCACGAGGTGGAAAAAGGGGAGAG-----CCACTGGAGCGCTGCTACTCT	273
Db	74	AspAlagly-AspleuaspileGluAspMetLysAlaGlnMetAsnCysPheTyrl	93
Qy	274	GAGGCGCCTGGAGGTTTCGTCTAGTACTACCGCGCGAGGAGACATGGCTTACTGTCT	333
Db	93	uLysAlaLeuAspGlyPheValMetValLeuThrAspAspGlyAspMetIleTyrlleSe	113
Qy	334	GGAAAATGTCAAGACACCTGGCGCTCAGTCAGTGGACCTCTGTTCCTCCCTCCCTGATA	393
Db	113	rAspAsnValAsnLysTyrmMetGlyLeuthrGln-----	124
Qy	394	CATAACCCCACTCCTGGTAGCAAAATTCTCTCTGGAGCTCATTTGGACAGACTATCTTGTAT	453









```
QY 79 GACGCGCCGACGCGCGCAGCCAGGACGAGGTGCTGTACCACTGGCGCACACT 138
DB 21 AspaAlaAargCysArgSerLysGluThrGluValPheTyrGluLeuAlaHisGlu 40
QY 139 CTGCGCCCTTGGCGCGCGGTAGCGGCACCTGGACAAAGCCCTCCATCATGCGCCCTACA 198
DB 41 LeuProLeuProHisSerValSerHisLeuAspLysAlaSerIleMetArgLeuAla 60
QY 199 ATCAGCTACCTGGCATGCACCGCCTC-----TGGCAGCAGGTGAAAGG 246
DB 61 IleSerPheLeuArgThrHisLysLeuLeuSerValCysSer-GluAsnGluSerG1 80
QY 247 GGGAGAG-----CCACTGGACGCGCTGTACTCTGAAGCGCCTGGAGGTTTCGTCAT 297
DB 80 uAlaGluAlaAspGlnGlnMetAspAsnLeuTyrLeuLysAlaLeuGluGlyPheIleAl 100
QY 298 GGTACTCACCAGCGCGGAGGACATGCTTACCTGCGGAAATGTCAGCAAGACCTGGG 357
DB 100 aValValThrGlnAspGlyAspMetIlePheLeuSerGluAsnIleSerLysPheMetG1 120
QY 358 CCTCAGTCAGTGGACCTCTGTTCTCTCCTCGATACATAACCCACTCTCGTTACCAAT 417
DB 120 yLeuThrGln----- 123
QY 418 TTCTCTCTGGAGCTCATTTGGACAGCATATCTTTGATTTATCCATCCTCTGTGACCAAGAG 477
DB 124 ---Val--GluLeuThrGlyHisSerIlePheAspPheThrHisProCysAspHisGlu 141
QY 478 GAACCTCAAGACGCCCTGACCCCGAG-----CCGAACTGTCAAAGAAGAGCTGGAA 531
DB 142 GluIleArgGluAsnLeuSerLeuLysAsnGlySerGlyPheGlyLysLysSerLysAsp 161
QY 532 GCCCAACAGAGCGCCATTTTCCCTGCGAATGAAGACACGCTCACACAGAGGCGC 591
DB 162 MetSerThrGluArgAspPhePheMetArgMetLysCysThrValThrAsnArgGlyArg 181
QY 592 ACCTCAACCTCAAAGCGCCACCTGGAGAGGTGCTGCTCAGGACATATGAGGGCC 651
DB 182 ThrValAsnLeuLysSerAlaThrTrpLysValLeuHisCysThrGlyGlnValLysVal 201
QY 652 TACAAGCCCC--CCTGCACAGACTTCCCTCGCGGGAGCCCTCGCTCCGAGCCTCCCTG 708
DB 202 TyrAsnAsnCysProProHisAsnSerLeuCysGly-----TyrLysGluProLeu 219
QY 709 CAATGCTGTGCTTATCTGTGAAGCATCCCGCAGCTCCCTTCCAC 756
DB 220 SerCysLeuIleIleMetCysGluProIle---GlnHisProSerHis 234

RESULT 15
US-09-374-454-4
; Sequence 4, Application US/09374454
; Patent No. 6395548
; GENERAL INFORMATION:
; APPLICANT: Lee, Mu-En
; APPLICANT: Maemura, Koji
; APPLICANT: Hsieh, Chung-Ming
; TITLE OF INVENTION: METHODS OF MODULATING OF ANGIOGENESIS
; FILE REFERENCE: 05433/037001
; CURRENT APPLICATION NUMBER: US/09/374,454
; CURRENT FILING DATE: 1999-08-13
; EARLIER APPLICATION NUMBER: US 60/096,515
; EARLIER FILING DATE: 1998-08-14
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 485
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-374-454-4
```

Alignment Scores: 2.9e+48 Length: 485  
Pred. No.:

```
Score: 571.50 Matches: 129
Percent Similarity: 64.34% Conservative: 37
Best Local Similarity: 50.00% Mismatches: 58
Query Match: 28.43% Indels: 34
DB: 4 Gaps: 7

US-09-896-791B-2 (1-1100) x US-09-374-454-4 (1-485)

QY 19 ATCGCGTTGGGCTGCGAGCGGTGAGTGCACACCGAGCTGCGGAGGAGAGTGC GCG 78
DB 1 MetThrAlaAspLysGluLysLysArgSerSerGluArgArgLysGluLysSerArg 20
QY 79 GACGCGCCGCGCAGCGCGGCGCAGCCAGGACGAGGTGCTGTACCACTGGCGCACACT 138
DB 21 AspaAlaAargCysArgSerLysGluThrGluValPheTyrGluLeuAlaHisGlu 40
QY 139 CTGCGCCCTTGGCGCGCGGTAGCGGCACCTGGACAAAGCCCTCCATCATGCGCCCTACA 198
DB 41 LeuProLeuProHisSerValSerHisLeuAspLysAlaSerIleMetArgLeuGlu 60
QY 199 ATCAGCTACCTGGCATGCACCGCCTC-----TGGCAGCAGGTGAAAGG 246
DB 61 IleSerPheLeuArgThrHisLysLeuLeuSerValCysSer-GluAsnGluSerG1 80
QY 247 GGGAGAG-----CCACTGGACGCGCTGTACTGTGAAGCGCCTGGAGGTTTCGTCAT 297
DB 80 uAlaGluAlaAspGlnGlnMetAspAsnLeuTyrLeuLysAlaLeuGluGlyPheIleAl 100
QY 298 GGTACTCACCAGCGCGGAGGACATGCTTACCTGCGGAAATGTCAGCAAGACCTGGG 357
DB 100 aValValThrGlnAspGlyAspMetIlePheLeuSerGluAsnIleSerLysPheMetG1 120
QY 358 CCTCAGTCAGTGGACCTCTGTTCTCTCCTCGATACATAACCCACTCTCGTTACCAAT 417
DB 120 yLeuThrGln----- 123
QY 418 TTCTCTCTGGAGCTCATTTGGACAGCATATCTTTGATTTATCCATCCTCTGTGACCAAGAG 477
DB 124 ---Val--GluLeuThrGlyHisSerIlePheAspPheThrHisProCysAspHisGlu 141
QY 478 GAACCTCAAGACGCCCTGACCCCGAG-----CCGAACTGTCAAAGAAGAGCTGGAA 531
DB 142 GluIleArgGluAsnLeuSerLeuLysAsnGlySerGlyPheGlyLysLysSerLysAsp 161
QY 532 GCCCAACAGAGCGCCATTTTCCCTGCGAATGAAGACACGCTCACACAGAGGCGC 591
DB 162 MetSerThrGluArgAspPhePheMetArgMetLysCysThrValThrAsnArgGlyArg 181
QY 592 ACCTCAACCTCAAAGCGCCACCTGGAGAGGTGCTGCTCAGGACATATGAGGGCC 651
DB 182 ThrValAsnLeuLysSerAlaThrTrpLysValLeuHisCysThrGlyGlnValLysVal 201
QY 652 TACAAGCCCC--CCTGCACAGACTTCCCTCGCGGGAGCCCTCGCTCCGAGCCTCCCTG 708
DB 202 TyrAsnAsnCysProProHisAsnSerLeuCysGly-----TyrLysGluProLeu 219
QY 709 CAATGCTGTGCTTATCTGTGAAGCATCCCGCAGCTCCCTTCCAC 756
DB 220 SerCysLeuIleIleMetCysGluProIle---GlnHisProSerHis 234

Search completed: December 2, 2002, 20:41:46
Job time : 32 secs
```

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM nucleic - protein search, using frame\_plus\_n2p model  
Run on: December 2, 2002, 20:39:21 : Search time 40 Seconds  
(without alignments)  
875.831 Million cell updates/sec

Title: US-09-896-791B-2  
Perfect score: 2010  
Sequence: 1 gaattcgacagggccat.....aaaaaaaaaacatcgggcgc 1100

Scoring table: BLOSUM62

Xgapop 10.0, Xgapext 0.5  
Ygapop 10.0, Ygapext 0.5  
Fgapop 6.0, Fgapext 7.0  
Delop 6.0, Delext 7.0

Searched: 102317 seqs, 15924203 residues

Total number of hits satisfying chosen parameters: 204634

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame\_n2p.model -DEV=xlpl  
-Q=/cgn2\_1/USProo1/US09896791/runat\_02122002\_152700\_5721/app\_query.fasta\_1.1287  
-DB=PublishedApplications\_AA -QPMF=fastan -SUFFIX=rapb -MNMATCH=0.1  
-LOOPCL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blomum62  
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR SCORE=pct -THR MAX=100  
-THR\_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0  
-MAXLEN=2000000000 -USER=US09896791@cgn\_1.1.5 @runat\_02122002\_152700\_5721  
-NCPU=6 -ICPU=3 -NO\_XLPXY -NO\_MMAP -LARGEQUERY -NEG\_SCORES=0 -WAIT -LONGLOG  
-DEV\_TIMEOUT=120 -WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Published Applications\_AA:

- 1: /cgn2\_6/ptodata/1/pubpaa/US08\_NEW\_PUB.pap:
- 2: /cgn2\_6/ptodata/1/pubpaa/PCT\_NEW\_PUB.pap:
- 3: /cgn2\_6/ptodata/1/pubpaa/US06\_NEW\_PUB.pap:
- 4: /cgn2\_6/ptodata/1/pubpaa/US06\_PUBCOMB.pap:
- 5: /cgn2\_6/ptodata/1/pubpaa/US07\_NEW\_PUB.pap:
- 6: /cgn2\_6/ptodata/1/pubpaa/US07\_PUBCOMB.pap:
- 7: /cgn2\_6/ptodata/1/pubpaa/PCTUS\_PUBCOMB.pap:
- 8: /cgn2\_6/ptodata/1/pubpaa/US08\_PUBCOMB.pap:
- 9: /cgn2\_6/ptodata/1/pubpaa/US09\_NEW\_PUB.pap:
- 10: /cgn2\_6/ptodata/1/pubpaa/US09\_PUBCOMB.pap:
- 11: /cgn2\_6/ptodata/1/pubpaa/US10\_NEW\_PUB.pap:
- 12: /cgn2\_6/ptodata/1/pubpaa/US10\_PUBCOMB.pap:
- 13: /cgn2\_6/ptodata/1/pubpaa/US60\_NEW\_PUB.pap:
- 14: /cgn2\_6/ptodata/1/pubpaa/US60\_PUBCOMB.pap:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description
1	1636	81.4	307	9	US-09-896-791B-3
2	606.5	30.2	826	9	US-09-736-457-330
3	606.5	30.2	826	9	US-09-902-941-330
4	606.5	30.2	826	10	US-09-922-958-4

5	606.5	30.2	826	10	US-09-833-790-235
6	606.5	30.2	826	12	US-10-028-158-23
7	311	15.5	570	10	US-09-923-684-4
8	311	15.5	667	10	US-09-923-684-3
9	222.5	11.1	99	10	US-09-925-297-863
10	192	9.6	333	10	US-09-925-302-659
11	162	8.1	591	10	US-09-764-864-1139
12	153	7.6	53	10	US-09-864-761-41589
13	144.5	7.2	530	10	US-09-800-729-112
14	134	6.7	399	9	US-09-764-868-850
15	132.5	6.7	200	10	US-09-764-864-953
16	128.5	6.4	1464	10	US-09-842-256-2
17	127.5	6.3	223	10	US-09-925-301-1280
18	126	6.3	181	10	US-09-764-846-264
19	123	6.3	178	9	US-09-764-868-650
20	120.5	6.0	292	10	US-09-764-853-604
21	119.5	5.9	315	10	US-09-764-853-793
22	119	5.9	219	12	US-10-001-843-143
23	116.5	5.8	640	12	US-10-028-780-2
24	115	5.9	60	10	US-09-764-846-178
25	115	5.7	359	10	US-09-864-761-36539
26	115	5.7	846	10	US-09-858-684A-3
27	114.5	5.7	133	10	US-09-925-301-1086
28	113	5.6	201	10	US-09-864-864-1561
29	111.5	5.5	881	10	US-09-816-860A-2
30	111	5.5	51	10	US-09-764-846-151
31	111	5.5	154	10	US-09-825-301-27
32	111	5.5	154	10	US-09-604-287A-471
33	111	5.5	154	12	US-10-007-805-471
34	110.5	5.5	677	10	US-09-925-300-1626
35	110.5	5.5	1350	10	US-09-893-238-17
36	109	5.4	196	10	US-09-764-846-251
37	108.5	5.4	645	10	US-09-925-300-1453
38	107.5	5.3	150	10	US-09-764-846-225
39	107.5	5.3	388	10	US-09-925-300-1618
40	107.5	5.3	1317	10	US-09-963-896-7
41	107	5.4	80	10	US-09-764-846-158
42	106.5	5.3	1429	10	US-09-996-617-2
43	106.5	5.3	1429	10	US-09-931-071-2
44	106	5.4	50	10	US-09-764-846-186
45	106	5.4	60	10	US-09-764-846-187

ALIGNMENTS

RESULT 1  
US-09-896-791B-3  
; Sequence 3, Application US/09896791B  
; Patent No. US20020165140A1  
; GENERAL INFORMATION:  
; APPLICANT: Berkenstam, Anders  
; APPLICANT: Bertilsson, Gran  
; APPLICANT: Poellinger, Lorenz  
; TITLE OF INVENTION: SCREENING METHODS  
; FILE REFERENCE: 13425-040001  
; CURRENT APPLICATION NUMBER: US/09/896,791B  
; CURRENT FILING DATE: 2001-06-29  
; PRIOR APPLICATION NUMBER: 60/217,570  
; PRIOR FILING DATE: 2000-07-12  
; PRIOR APPLICATION NUMBER: SE 0002551-0  
; PRIOR FILING DATE: 2000-07-06  
; NUMBER OF SEQ ID NOS: 3  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 3  
; LENGTH: 307  
; TYPE: PRT  
; ORGANISM: Mus musculus  
US-09-896-791B-3

Alignment Scores: 3.93e-120 Length: 307  
Pred. No.: 1636.00 Matches: 307  
Score: 100.00%  
Percent Similarity: 100.00%  
Conservative: 0

Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 81.39% Indels: 0  
DB: 9 Gaps: 0

US-09-896-791B-2 (1-1100) x US-09-896-791B-3 (1-307)

```
QY 19 ATGCGTTGGGCTGCGCGCGTGAAGTGAACCGAGCTGCGGAGGAGAGTGC GGG 78
Db 1 MetAlaLeuGlyLeuGlnArgValArgSerAsnThrGluLeuArgLysGluLysSerArg 20
QY 79 GACCGCGCCGCGCAGCGCGCGCAGCGAGGAGGAGTGTGTACCAAGTGGCGCACACT 138
Db 21 AspAlaAlaArgSerArgSerGlnGlnThrGluValLeuLeuAlaHisThr 40
QY 139 CTGCGCTTTGCGCGCGCTGAGCGCGCAGTGGACAGGCTTCATCATCGCGCTCACA 198
Db 41 LeuProPheAlaArgGlyValSerAlaHisLeuAspLysAlaSerIleMetArgLeuThr 60
QY 199 ATCAGCTACCTGCGCATGCCCGCTCTGCGCAGCAGGTGGAAAGGGGAGAGCCACT 258
Db 61 IleSerTyrLeuArgMetHisArgLeuCysAlaAlaGlyGlyLysArgGlyArgAlaThr 80
QY 259 GGACCGCTGTACTGAAGCGCTGGAGGTTTCGTGTCATGTTACTACCGCCGAGGAGA 318
Db 81 GlyArgLeuLeuProGluGlyProGlyGlyPheArgHisGlyThrHisArgGlyArg 100
QY 319 CATGGCTTACCTGTCGGAATGTCAGCAGCACCTGGCGCTCAGTCAGTGGACCTCTGT 378
Db 101 HisGlyLeuProValGlyLysCysGlnGlnAlaProGlyProGlnSerValAspLeuCys 120
QY 379 TCCTCTCTCCCTGATACATAACCCACCTCCCTGGTACCAATTTCTCTCTGGAGCTCATTTGA 438
Db 121 SerSerSerLeuIleHisAsnProThrProGlyThrAsnPheSerLeuGluLeuIleGly 140
QY 439 CACAGTATCTTTGATTTTATCCATCCCTGTGACCAAGAGGAATTCGAAGCGCCCTGAAC 498
Db 141 HisSerIlePheAspPheIleHisProCysAspGlnGluLeuGlnAspAlaLeuThr 160
QY 499 CCAGGCGCGAACCTGTCAAAGAAGAGCTGGAAGCCCAACAGAGCGCCACTTTCCCTG 558
Db 161 ProArgProAsnLeuSerLysLysLysLeuGluAlaProThrGluArgHisPheSerLeu 180
QY 559 CGAATGAGAGCAGCTCACCAGCAGAGGCGCACGCTCAACCTCAAAGCGGCCACCTGG 618
Db 181 ArgMetLysSerThrLeuThrSerArgGlyArgThrLeuAsnLeuLysAlaAlaThrTrp 200
QY 619 AAGTGCTGCTACGCTCAGACATATAGGCGCTTACAGCCCTTACAGCCCTGACAGACTTCCCT 678
Db 201 LysValLeuHisCysSerGlyHisMetArgAlaTyrLysProAlaGlnThrSerPro 220
QY 679 GCCGGGAGCCCTCGCTCCGAGCCTCCCTGCAATGCTGCTGTATCTCTGAGCGCATC 738
Db 221 AlaGlySerProArgSerGluProProLeuGlnCysLeuValLeuIleCysGluAlaIle 240
QY 739 CCCAGCTCCCTTCCACGATGTTGCTACTCTGGGTCTTCCACAGGAGAGACTCCCATC 798
Db 241 ProGlnLeuProPheHisAspGlyAlaThrLeuGlyLeuProGlnGluLysThrProIle 260
QY 799 TCTACCTTATACCCCTCTTTGAGAGGCACTACTTTGCTTGTTCAGAGAGTGGCTGTT 858
Db 261 SerThrLeuPheThrProLeuTrpLysAlaLeuLeuCysLeuValLysArgTrpProVal 280
QY 859 CAGTGCTACAGGGGAAAGGACTGAATCTCTCTCCCTCATGGGTGTTGTGGGCCCTT 918
Db 281 GlnValLeuGlnGlyLysGlyThrGluSerSerLeuProSerTrpValLeuTrpAlaLeu 300
QY 919 AACCGGAAAAATGTCCTGGC 939
Db 301 AsnArgLysAsnCysProGly 307
```

RESULT 2

US-09-736-457-330

; Sequence 330, Application US/09736457

; Patent No. US20020168637A1  
; GENERAL INFORMATION:  
; APPLICANT: Wang, Tongtong  
; APPLICANT: Bangur, Chaitanya S.  
; APPLICANT: Lodes, Michael A.  
; APPLICANT: Fanger, Gary  
; APPLICANT: Vedvick, Tom  
; APPLICANT: Carter, Darrick  
; APPLICANT: Retter, Marc  
; APPLICANT: Mannion, Jane  
; APPLICANT: Fan, Liqun  
; APPLICANT: Wang, Aijun  
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND  
; FILE REFERENCE: 210121.478C15  
; CURRENT APPLICATION NUMBER: US/09/736,457  
; CURRENT FILING DATE: 2000-12-13  
; NUMBER OF SEQ ID NOS: 1864  
; SOFTWARE: FASTSEQ for Windows Version 3.0  
; SEQ ID NO 330  
; LENGTH: 826  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; US-09-736-457-330

Alignment Scores:

Pred. No.: 7,65e-40 Length: 826  
Score: 606.50 Matches: 129  
Percent Similarity: 68.07% Conservative: 33  
Best Local Similarity: 54.20% Mismatches: 48  
Query Match: 30.17% Indels: 28  
DB: 9 Gaps: 3

US-09-896-791B-2 (1-1100) x US-09-736-457-330 (1-826)

```
QY 49 AACACCGAGCTGCGGAGGAGAAAGTTCGCGGAGCGCGCGCGAGCGCGCGAGCGGCGCAGGAG 108
Db 14 SerSerGluArgArgLysGluLysSerArgAspAlaAlaArgSerArgArgSerLysGlu 33
QY 109 ACGAGGTGCTGTACAGCTGGCGGCGACACTCTGCTCCCTTTTCGCGCGCGCTCAGCGCGCAC 168
Db 34 SerGluValPheTyrGluLeuAlaHisGlnLeuProLeuProHisAsnValSerSerHis 53
QY 169 CTGGAAGCGCTTCATCATCGCGCTCACAAATCAGCTACCTGCGCATGCGCACCGCTCTGC 228
Db 54 LeuAspLysAlaSerValMetArgLeuThrIleSerTyrLeuArgValArgLysLeuLeu 73
QY 229 GCAGCAGGTGGAAAAAGGGGAGAG-----CCACTGGACGCGCTGTACCT 273
Db 74 AspAlaGly-AspLeuAspIleGluAspMetLysAlaGlnMetAsnCysPheTyrLe 93
QY 274 GAAGGCCCTGGAGGTTTCGTATGTTACTCACCAGCGGAGGAGACATGCTTACCTGTC 333
Db 93 uLysAlaLeuAspGlyPheValMetValLeuThrAspAspGlyAspMetIleTyrIleSe 113
QY 334 GGAATATGTCAAGAACACCTGCGCTCAGTCAAGTGGACCTCTCTCTCTCTCTCTCTCTCT 393
Db 113 rAspAsnValAsnLysTyrMetGlyLeuThrGln----- 124
QY 394 CATAACCCCACTCCTGTGTACCAATTTCTCTCTGTGAGCTCATTTGGACACAGTATCTTTGAT 453
Db 125 -----Phe--GluLeuThrGlyHisSerValPheAsp 134
QY 454 TTTTATCCATCCCTGTGACCAAGAGGAACACTTCAAGACGCGCTGACCCCGGAGCGGAGCGGAG 513
Db 135 PheThrHisProCysAspHisGluGluMetArgGluMetLeuThrHisArgAsnGlyLeu 154
QY 514 TCAAAGAAAGAGCTGGAAGCCCAACAGAGCGCACCTTTTCCTCTGCGCAATGAAGAGACAG 573
Db 155 valLysGlyLysGluGlnAsnThrGlnArgSerPhePheLeuArgMetLysCysThr 174
QY 574 CTCACGAGCAGAGGCGCGACGCTCAACCTCAAAAGCGGCCACCTGGAAGGTGCTGCACTGC 633
|||||
```

Db 175 LeuThrSerArglyArgThrMetAsnIleLysSerAlaThrTrpLysValLeuHisCys 194  
QY 634 TCAGGACATATGAGGCGCTACAAGCCCTGACACAGACTTCCCTCGCGGAGCCCTCGC 693  
Db 195 ThrGlyHisIleHisVallyrAspThrAsnSerAsnGlnProGlnCysGly-----Tyr 212  
QY 694 TCCGAGCCCTCCCTGCAATGCCCTGGTGTATCTGTGAAGCCATCCCC 741  
Db 213 LysLysProPromethrCysLeuValLeuIleCysGluProIlePro 228  
RESULT 3  
US-09-902-941-330  
; Sequence 330, Application US/09902941  
; Patent No. US20020172952A1  
; GENERAL INFORMATION:  
; APPLICANT: Henderson, Robert A.  
; APPLICANT: Wang, Tongtong  
; APPLICANT: Watanabe, Yoshihiro  
; APPLICANT: Johnson, Jeffrey C.  
; APPLICANT: Retter, Marc W.  
; APPLICANT: Marnerakis, Margarita  
; APPLICANT: Carter, Darrick  
; APPLICANT: Fanger, Gary R.  
; APPLICANT: Vedwick, Thomas S.  
; APPLICANT: Bangur, Chaitanya S.  
; APPLICANT: McNabb, Andria  
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY  
; TITLE OF INVENTION: AND DIAGNOSIS OF LUNG CANCER  
; FILE REFERENCE: 210121.478C17  
; CURRENT APPLICATION NUMBER: US/09/902,941  
; CURRENT FILING DATE: 2001-07-10  
; NUMBER OF SEQ ID NOS: 2002  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 330  
; LENGTH: 826  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-902-941-330  
Alignment Scores:  
Pred. No.: 7.65e-40 Length: 826  
Score: 606.50 Matches: 129  
Percent Similarity: 68.07% Conservative: 33  
Best Local Similarity: 54.20% Mismatches: 48  
Query Match: 30.17% Indels: 28  
DB: 9 Gaps: 3  
US-09-896-791B-2 (1-1100) x US-09-902-941-330 (1-826)  
QY 49 AACACCGAGCTGCGGAAGAGAGTCCGGGACGCGCGCGCGGCGGCGGCGGCGGAG 108  
Db 14 SerSerGluArgArgLysGluLysSerArgAspAlaAlaArgSerArgSerLysGlu 33  
QY 109 ACGAGGTGCTGTACAGCTGGCGGCACACTCTGCCCTTTTCCGCGCGCGCTCAGCGCGCAC 168  
Db 34 SerGluValPheTyrGluLeuAlaHisGlnLeuProLeuProHisAsnValSerSerHis 53  
QY 169 CTGGACAAGCCCTCCATCATGCGCTCACAATCAGTACTGCGCATGACCGCCCTCTGC 228  
Db 54 LeuAspLysAlaSerValMetArgLeuThrIleSerTyrLeuArgValArgLysLeuLeu 73  
QY 229 GCAGAGGTGGAAAAGGGGAGAG-----CCACTGGAGCGCTGTACCT 273  
Db 74 AspAlaGly-AspLeuAspIleGluAspMetLysAlaGlnMetCAsnCysPheTyrLe 93  
QY 274 GAAGGCCCTGGAGGTTTTCGTATGTTACTACCGCGGAGGAGACATGGCTTACCTGTG 333  
Db 93 uLysAlaLeuAspGlyPheValMetValLeuThrAspAspGlyAspMetIleTyrIleSe 113  
QY 334 GGAAATGTCAGCAGCACCTGGGCCCTCAGTCAGTGGAGCTCTGTTCCTCCTCCCTGATA 393  
Db 113 rAspAsnValAsnLysTyrMetGlyLeuThrGln----- 124

QY 394 CATAACCCCACTCCTGCTACCAATTTCTCTCTGGAGCTCATTTGGACACAGTATCTTTGAT 453  
Db 125 -----Phe--GluLeuThrGlyHisSerValPheAsp 134  
QY 454 TTTATCATCCCTGTGACCAAGAGAACTTCAAGAGCCCTGACCCCGGAGCCGACCTG 513  
Db 135 PheThrHisProcysAspHisGluGluMetArgGluMetLeuThrHisArgAsnGlyLeu 154  
QY 514 TCAAGAAGAAGCTGGAAGCCCCCAACAGAGCGCCACTTTTCCCTGCGAATGAAGAGCAG 573  
Db 155 ValLysLysGlyLysGluGlnAsnThrGlnArgSerPhePheLeuArgMetLysCysThr 174  
QY 574 CTCACGAGGAGGCGGCACGCTCAACCTCAAAGCGGCCACCTGGAGGTGCTGCCTGC 633  
Db 175 LeuThrSerArgGlyArgThrMetAsnIleLysSerAlaThrTrpLysValLeuHisCys 194  
QY 634 TCAGGACATATGAGGCGCTACAAGCCCTGTCAGACACTTCCCTCGCGGAGCCCTCGC 693  
Db 195 ThrGlyHisIleHisValTyrAspThrAsnSerAsnGlnProGlnCysGly-----Tyr 212  
QY 694 TCCGAGCCCTCCCTGCAATGCCCTGGTGTATCTGTGAAGCCATCCCC 741  
Db 213 LysLysProPromethrCysLeuValLeuIleCysGluProIlePro 228  
RESULT 4  
US-09-922-958-4  
; Sequence 4, Application US/09922958  
; Patent No. US20020048794A1  
; GENERAL INFORMATION:  
; APPLICANT: POELLINGER, Lorenz  
; APPLICANT: PEREIRA, Teresa  
; APPLICANT: RUAS, Jorge  
; TITLE OF INVENTION: MECHANISM OF CONDITIONAL REGULATION OF THE HYPOXIA-INDUCIBLE F  
; TITLE OF INVENTION: THE VON HIPPEL-LINDAU TUMOR SUPPRESSOR PROTEIN  
; FILE REFERENCE: 3743/49008  
; CURRENT APPLICATION NUMBER: US/09/922,958  
; CURRENT FILING DATE: 2001-08-07  
; PRIOR APPLICATION NUMBER: US 60/223,480  
; PRIOR FILING DATE: 2000-08-07  
; NUMBER OF SEQ ID NOS: 7  
; SOFTWARE: Patentin version 3.0  
; SEQ ID NO 4  
; LENGTH: 826  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-922-958-4  
Alignment Scores:  
Pred. No.: 7.65e-40 Length: 826  
Score: 606.50 Matches: 129  
Percent Similarity: 68.07% Conservative: 33  
Best Local Similarity: 54.20% Mismatches: 48  
Query Match: 30.17% Indels: 28  
DB: 10 Gaps: 3  
US-09-896-791B-2 (1-1100) x US-09-922-958-4 (1-826)  
QY 49 AACACCGAGCTGCGGAAGAGAGTCCGGGACGCGCGCGCGGCGGCGGCGGAG 108  
Db 14 SerSerGluArgArgLysGluLysSerArgAspAlaAlaArgSerArgSerLysGlu 33  
QY 109 ACGAGGTGCTGTACAGCTGGCGGCACACTCTGCCCTTTTCCGCGCGCGCTCAGCGCGCAC 168  
Db 34 SerGluValPheTyrGluLeuAlaHisGlnLeuProLeuProHisAsnValSerSerHis 53  
QY 169 CTGGACAAGCCCTCCATCATGCGCTCACAATCAGTACTGCGCATGACCGCCCTCTGC 228  
Db 54 LeuAspLysAlaSerValMetArgLeuThrIleSerTyrLeuArgValArgLysLeuLeu 73  
QY 229 GCAGAGGTGGAAAAGGGGAGAG-----CCACTGGAGCGCTGTACCT 273  
Db 74 AspAlaGly-AspLeuAspIleGluAspMetLysAlaGlnMetCAsnCysPheTyrLe 93

QY 274 GAAGGCGCTCGAGGGTTTCCTCATGGTACATCAACGCGGAGGAGACATGCTTACCTGTC 333  
Db 93 uLysAlaLeuAspGlyPheValMetValLeuThrAspGlyAspMetIleTyrlleSe 113  
QY 334 GGAATATGTCAGCAAGCACCTGGGCTCAGTCACTGAGTCACTGCTCTCTCCCTGATA 393  
Db 113 rAspAsnValAsnLysTyrrMetGlyLeuThrGln----- 124  
QY 394 CATAACCCCACTCTGCTGACCAATTTCTCTGAGCTCATTTGGACACAGTATCTTTGAT 453  
Db 125 -----Phe--GlulLeuThrGlyHisSerValPheAsp 134  
QY 454 TTTATCCATCCCTGTGACCAAGAGGAACTTCAAGACGCCCTGACCCCGAGGCGCAACCTG 513  
Db 135 PheThrHisProCysAspHisGluMetArgGluMetLeuThrHisArgAsnGlyLeu 154  
QY 514 TCAAGAAGAGCTGGAGGCCCAACAGAGCGGCACCTTTCCCTGCGAATGAAGAGCAGC 573  
Db 155 ValLysLysGlyLysGluGlnAsnThrGlnArgSerPhePheLeuArgMetLysCysThr 174  
QY 574 CTCACCAAGAGCGGCGCACGCTCAAGCGGCCCACTGGAAGGTGCTGCACTGC 633  
Db 175 LeuThrSerArgGlyArgThrMetAsnIleLysSerAlaThrTrpLysValLeuHisCys 194  
QY 634 TCAGGACATATAGGCGCTACAAGCCCCCTGCACAGACTTCCCTGCGGAGGCCCTCGC 693  
Db 195 ThrGlyHisIleHisValTyrrAspThrAsnSerAsnGlnProGlnCysGly-----Tyr 212  
QY 694 TCCGAGCTCCCTGCAATCCCTGGTGGTCTTATCTGTGAAGCCATCCCC 741  
Db 213 LysLysProMetThrCysLeuValLeuIleCysGluProIlePro 228

## RESULT 5

US-09-833-790-235  
; Sequence 235, Application US/09833790  
; Patent No. US20020068288A1  
; GENERAL INFORMATION:  
; APPLICANT: Lodes, Michael J.  
; APPLICANT: Wang, Tongtong  
; APPLICANT: Secrist, Heather  
; APPLICANT: Mohamath, Raodoh  
; APPLICANT: Indirias, Carol Y.  
; APPLICANT: Fan, Liqun  
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY  
; FILE OF INVENTION: AND DIAGNOSIS OF LUNG CANCER  
; FILE REFERENCE: 210121-512  
; CURRENT APPLICATION NUMBER: US/09/833,790  
; CURRENT FILING DATE: 2001-04-11  
; NUMBER OF SEQ ID NOS: 440  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 235  
; LENGTH: 826  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-833-790-235

Alignment Scores:  
Pred. No.: 7,65e-40 Length: 826  
Score: 606.50 Matches: 129  
Percent Similarity: 68.07% Conservative: 33  
Best Local Similarity: 54.20% Mismatches: 48  
Query Match: 30.17% Indels: 28  
DB: Gaps: 3

US-09-896-791B-2 (1-1100) x US-09-833-790-235 (1-826)

QY 49 AACACGAGTGGGAGGAGAGTCCGCGGACGCGCGCCGCGGAGGAGGAGGAGGAGGAG 108  
Db 14 SerSerGluArgLysGluLysSerArgAspAlaAlaArgSerArgSerLysGlu 33  
QY 109 ACGAGGTGCTACCAAGTGGCGCACACTCTGCCCTTTGCGGCGCGCTCAGCGCGCAC 168  
Db 34 SerGluValPheTyrrGlyLeuAlaHisGlnLeuProLeuProHisAsnValSerSerHis 53

17-6

QY 169 CTGACAAGGCTCCATCATCGCCTCACATCAGCTACCTGGCGCATCGCCGCTCTGC 228  
Db 54 LeuAspLysAlaSerValMetArgLeuThrIleSerTyrrLeuArgValArgLysLeuLe 73  
QY 229 GCAGCAGTGGAAAAGGGGAGAG-----CCACTGGACGCTGCTACCT 273  
Db 74 AspAlaGly-AspLeuAspIleGluAspAspMetLysAlaGlnMetAsnCysPheTyrrLe 93  
QY 274 GAAGGCGCTGGAGGTTCCTCATGCTGCTACCTGAGGAGGAGGAGACATGCTTACCTGTC 333  
Db 93 uLysAlaLeuAspGlyPheValMetValLeuThrAspGlyAspMetIleTyrlleSe 113  
QY 334 GGAATATGTCAGCAAGCACCTGGGCTCAGTCACTGAGTCACTGCTCTCTCCCTGATA 393  
Db 113 rAspAsnValAsnLysTyrrMetGlyLeuThrGln----- 124  
QY 394 CATAACCCCACTCTGCTGACCAATTTCTCTGAGCTCATTTGGACACAGTATCTTTGAT 453  
Db 125 -----Phe--GlulLeuThrGlyHisSerValPheAsp 134  
QY 454 TTTATCCATCCCTGTGACCAAGAGGAACTTCAAGACGCCCTGACCCCGAGGCGCAACCTG 513  
Db 135 PheThrHisProCysAspHisGluMetArgGluMetLeuThrHisArgAsnGlyLeu 154  
QY 514 TCAAGAAGAGCTGGAGGCCCAACAGAGCGGCACCTTTCCCTGCGAATGAAGAGCAGC 573  
Db 155 ValLysLysGlyLysGluGlnAsnThrGlnArgSerPhePheLeuArgMetLysCysThr 174  
QY 574 CTCACCAAGAGCGGCGCACGCTCAAGCGGCCCACTGGAAGGTGCTGCACTGC 633  
Db 175 LeuThrSerArgGlyArgThrMetAsnIleLysSerAlaThrTrpLysValLeuHisCys 194  
QY 634 TCAGGACATATAGGCGCTACAAGCCCCCTGCACAGACTTCCCTGCGGAGGCCCTCGC 693  
Db 195 ThrGlyHisIleHisValTyrrAspThrAsnSerAsnGlnProGlnCysGly-----Tyr 212  
QY 694 TCCGAGCTCCCTGCAATCCCTGGTGGTCTTATCTGTGAAGCCATCCCC 741  
Db 213 LysLysProMetThrCysLeuValLeuIleCysGluProIlePro 228

## RESULT 6

US-10-028-158-23  
; Sequence 23, Application US/10028158  
; Patent No. US20020110833A1  
; GENERAL INFORMATION:  
; APPLICANT: Caniggia, Isabella  
; APPLICANT: Post, Martin  
; APPLICANT: Lye, Stephen  
; TITLE OF INVENTION: METHODS TO DIAGNOSE A REQUIRED REGULATION OF  
; FILE OF INVENTION: TROPHOBLAST  
; FILE REFERENCE: 11757-38USWO  
; CURRENT APPLICATION NUMBER: US/10/028,158  
; CURRENT FILING DATE: 2001-12-20  
; PRIOR APPLICATION NUMBER: US/09/380,662  
; PRIOR FILING DATE: 1999-12-21  
; PRIOR APPLICATION NUMBER: PCT/CA98/00180  
; PRIOR FILING DATE: 1998-03-05  
; PRIOR APPLICATION NUMBER: US 60/039,919  
; PRIOR FILING DATE: 1997-03-07  
; NUMBER OF SEQ ID NOS: 24  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 23  
; LENGTH: 826  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-028-158-23

Alignment Scores:  
Pred. No.: 7,65e-40 Length: 826  
Score: 606.50 Matches: 129  
Percent Similarity: 68.07% Conservative: 33  
Best Local Similarity: 54.20% Mismatches: 48

Query Match:	30.17%	Indels:	28
DB:	12	Gaps:	3
US-09-896-791B-2 (1-1100) x US-10-028-158-23 (1-826)			
QY	49	AAACACGAGCTGCGGAAGGAGAAGTCCGGGACCGCGCGGCGGCGGACGAGCGGCGGCGGACGAGGAG	108
Db	14	SerSerGluArgArgLysGluLysSerArgAspAlaAlaArgSerArgArgSerLysGlu	33
QY	109	ACGGAGGTGCTGACACAGCTGGCGGCACACTGTCCTCTTTGCGCGCGGCTGCAGCGGCAC	168
Db	34	SerGluValPheTyrGluLeuAlaHisGlnLeuProLeuProHisasnValSerSerHis	53
QY	169	GTGCAAGAGCCCTCATCTGCGCCCTCACAAATCAGCTACTCGTCGCGATCGACCGCCTCTGC	228
Db	54	LeuaspLysAlaSerValMetArgLeuThrIleSerTyrLeuArgValArgLysLeuLeu	73
QY	229	GCAGCAGGTGGAAAAAGGGGAGAG-----CCACTGGAGCGCTGCTACT	273
Db	74	AspAlaGly-AspLeuAspIleGluAspMetLysAlaGlnMetAsnCysPheTyrLe	93
QY	274	GAAGGCCCTGGAGGTTTCGTGATGTTACTACCGCGCGGAGGAGACATGGCTTACCTGTC	333
Db	93	uLysAlaLeuAspGlyPheValMetValLeuThrAspAspGlyAspMetIleTyrIleSe	113
QY	334	GGAATAATGTCAGCAAGCACCTGGCGCTCAGTCAGTCAGTCGACCTCTGTTCCTCCTCGCTGATA	393
Db	113	raspasnValasnLysTyrMetGlyLeuThrGln-----	124
QY	394	CATAACCCCACTCCTGTTACCAATTCTCTCGAGCTCATTTGGACACAGTATCTTTGAT	453
Db	125	-----Phe--GluLeuThrGlyHisSerValPheasp	134
QY	454	TTTATCCATCCCTGTGACCAAGAGGAACCTTCAAGAGCCCTGACCCCGGCGGCAACCTG	513
Db	135	PheThrHisProCysAspHisGluGluMetArgGluMetLeuThrHisArgasnGlyLeu	154
QY	514	TCAAAGAGAAGCTGGAGCCCCAACAGAGCGCCACTTTTCCCTCGGAATGAAGAGCAGC	573
Db	155	ValLysGlyLysGluGlnAsnThrGlnArgSerPhePheLeuArgMetLysCysThr	174
QY	574	CTCACAGCAGAGGGCGCAGCTCAACCTCAAAGCGGCGCACCTGGAAGGTGCTGCACCTGC	633
Db	175	LeuThrSerArgLysArgThrMetasnIleLysSerAlaThrTrpLysValLeuHisCys	194
QY	634	TCAGGCATATGAGGGGCTTACAAGCCCCCTGTCACAGACTTCCCTCGCGGGAGCCCTCGC	693
Db	195	ThrGlyHisIleHisValTyrAspThrAsnSerAsnGlnProGlnCysGly-----Tyr	212
QY	694	TCCGAGCCTCCCTGCATGCGCTGGTGTATCTATCTGTGAAGCAATCCCC	741
Db	213	LysLysProMetThrCysLeuValLeuIleCysGluProIlePro	228

## RESULT 7

U-09-923-684-4

Sequence 4, Application US

Patent No. US20020081613A1

**GENERAL INFORMATION:**

**APPLICANT: Narayana**

**TITLE OF INVENTION:**

; FILE REFERENCE: 681

; CURRENT APPLICATION

; CURRENT FILING DATE

; NUMBER OF SEQ ID NO

; SOFTWARE: PatentIn

; SEQ ID NO 4

```

; LENGTH: 570

```

```

; TYPE: PRT

```

; ORGANISM: Homo sap

US-09-923-684-4

At present +

Alignment Scores:

Pred. No.:

Score:	311.00	Matches:	124
Percent Similarity:	42.38%	Conservative:	54
Best Local Similarity:	29.52%	Mismatches:	123
Query Match:	15.47%	Indels:	119
DB:	10	Gaps:	15

  

US-09-896~791B-2 (1-1100) x US-09-923-684-4 (1-570)

  

Qy	64	AAGGAGAAGTCGGCGGACGGCCGCCGCGCAGGCAGGCAGGAGCTGCTGTAC	123
Db			
Db	2	LysGluLysSerLysAsnAlaLaLysThrArgArgLysGluAsnGlyGluPheTyr	21
Qy	124	CAGCTGGCGCACACTCGCCCTTTTGCCGCGGGGTACGCCGCACTGGCAAGGCCTCC	183
Db		:::	
Db	22	GluLeuAlaLysLeuLeuProLeuProSerAlaIleThrSerGlnLeuAspLysAlaSer	41
Qy	184	ATCATGGCCCTCAACAATCAGCTACCTCGCATGCCCGCTCTGCCGACGAGT	237
Db			
Db	42	IleIleargLeuThrThrSerTyLeuLysMetArgAlaValPheProGluGlyLeuGly	61
Qy	238	-----GGAAAGGGGAGAGCC-	---
Db			----
Db	62	AspAlarppGlyGlnProSerArgAlaGlyProLeuAspGlyValAlaLysGluLeuGly	81
Qy	262	CGCCTGTACTCGTAAGCCCTGGAGGTTTTCGTCATGGTACTCACCGCCGAGGAGACAT	321
Db			
Db	82	SerHisLeu-LeuGlnThrLeuAspGlyPheValPheValValAlaSerAspGlyLysIl	101
Qy	322	GGCTTACCTGCGAAAATGTCACGACAGCCTGGCGCTCAGTCAGTGACCTCTGTCC	381
Db		:	
Db	101	eMetTyfileSerGluThrAlaSerValHisLeuGlyLeuSerGln-	116
Qy	382	TCCTCCCTGATACATAACCCCACTCCTGGTACCAAATTTCTCTGGAGCTCATTTGGACAC	441
Db		-----	--val--GluLeuThrGlyAsn
Db	117	-----	----
Qy	442	AGTATCTTTGATTATTCATCCCTGTGACCAAGAGAACCTTCAAGACGCCCTGACCCCC	501
Db			
Db	123	SeriletyrGluTyrlleHisProSerAspHisAspGluMetThrAlaValLeuThrAla	142
Qy	502	AGCGCGAACCCTGTCAAAGAGAAGCTG- --GAAGCCCACACAGAGCCCACTTTCCCTG	558
Db			
Db	143	HisGlnProLeuHisHisLeuLeuGlnGluTyrgluLeuGluArgSerPhePheLeu	162
Qy	559	CGAATGAAGAGCAGCGCTCACCGAGAGGCGCAGCGCTCAACCTCAAGGGCCACCTGG	618
Db			
Db	163	ArgMetLysCysValLeuAlaLys- ----ArgAsnAlaGlyLeuThrCysSerGlyTyrr	180
Qy	619	AAGTGCTGCTACTCTCAGGACATATGAGGCGCTACAAAGCCCTGCACAGACTCCCTCT	678
Db			
Db	181	LysValIleHisCysSerGlyTyrrLeuLysIleargGlnInTyrrMetLeuAspMetSerLeu	200
Qy	679	GCCGGGAGC-----	---CCTCGCTCCGAGCCCTCCCTCTG
Db			
Db	201	TyrAspSerCystyrGlnIleValGlyLeuValaValGlyGlnSerLeuProPro	219
Qy	709	CAATGCCTGGGTCTATCTGTGAAGGCATCCCCAGCTCCCTTCCACGATGCT	762
Db			
Db	220	-----SerAlaIlethrGluIleLysLeuTyrrSerAsnMetPhe	232
Qy	763	-----GCTACTCTGGGTCTT-	---
Db			
Db	233	MetPheArgAlaSerLeuAspLeuLysLeuIlePheLeuAspSerArgValThrGluVal	252
Qy	778	-----CCACAG-	-----GAGAAGACTCCCATC
Db			
Db	253	ThrGlytyrGluProGlnAspLeuIleGluLysThrLeuTyrrHisHisValHisGlyCys	272
Qy	802	ACCTTATTCACCCCTCTTTGGAAGGCACACTACTTTGCTTCTCAAGAGGTGGCCTGTT	858
Db		:::	
Db	273	AspValPheHisLeuArgTyrrAlaHisHisLeuLeuLeuValLysGlyGlnValThrThr	292
Qy	859	-----CAGGTGCTACAGGGGAAAGGG-	-----
Db		-----	-----

Alignment Scores:	
Pred. No.:	7.33e-17
Length:	570

```
Db 293 LysTyrTyrArgLeuLeuSerLysArgGlyGlyTrpValTrpValGlnSerTyrAlaThr 312
QY 880 -----ACTGAATCTCTCTCC-----TCATGGGTCTG 909
Db 313 ValValHisAsnSerArgSerArgProHisCysIleValSerValAsnTyrValLeu 332
QY 910 TGGGCCCTTAACCGGAAATTTGCTGCTAGGAGGAGTGAAGGACATGGCCACGCTA 969
Db 333 ThrGluIleGluTyrLysGlu-----LeuGlnLeu 342
QY 970 TCCTTACCCAGAAACCCCAAAATGCTCCAAACACCATAAAGACCTCTCTTGTGTAG 1029
Db 343 SerLeuGluGlnValSerThrAlaLysSerGlnAspSerTrpArgThrAlaLeuSerThr 362
QY 1030 GCACCAAGAAAAAATAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1083
Db 363 SerGlnGluThrArgLysLeuValLysProLysAsnThrLysMetLysThrLys 380
RESULT 8
US-09-923-684-3
; Sequence 3, Application US/09923684
; Patent No. US20020081613A1
; GENERAL INFORMATION:
; APPLICANT: Narayanan, Ramaswamy
; TITLE OF INVENTION: ASSOCIATION OF SIM2 WITH CANCER
; FILE REFERENCE: 6818-24
; CURRENT APPLICATION NUMBER: US/09/923,684
; CURRENT FILING DATE: 2001-09-17
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 3
; LENGTH: 667
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-923-684-3
Alignment Scores:
Pred. No.: 7.6e-17 Length: 667
Score: 311.00 Matches: 124
Percent Similarity: 42.38% Conservative: 54
Best Local Similarity: 29.52% Mismatches: 123
Query Match: 15.47% Indels: 119
DB: 10 Gaps: 15
US-09-896-791B-2 (1-1100) x US-09-923-684-3 (1-667)
QY 64 AAGGAGAAGTCGGGGAGCGGGCCGAGCGGGCGGAGCGGAGGAGGTGCTGTAC 123
Db 2 LysGluLysSerLysAsnAlaLysThrArgArgGluLysGluAsnGlyGluPheTyr 21
QY 124 CAGCTGGCGCACACTCTGCCCTTTGGCGGGGTCTAGCGCGCACCTGGACAGGCGCTCC 183
Db 22 GluLeuAlaLysLeuLeuProLeuProSerAlaIleThrSerGlnLeuAspLysAlaSer 41
QY 184 ATCATGGCCCTACAATCAGCTACCTACGATGCGCGCTCTCGCGAGCAGGT----- 237
Db 42 IleIleArgLeuThrThrSerTyrLeuLysMetArgAlaValPheProGluGlyLeuGly 61
QY 238 -----GGAAAGGGGGAGAGCC-----ACTGGA 261
Db 62 AspAlaTrpGlnProSerArgAlaGlyProLeuAspGlyValAlaLysGluLeuGly 81
QY 262 CGCTGTACTTGAAGCCCTGGAGGTTCGTGCTACTCACCCCGGAGGAGACAT 321
Db 82 SerHisLeu-LeuGlnThrLeuAspGlyPheValPheValAlaSerAspGlyLysII 101
QY 322 GGCTTACTGTGGAAATGTCAGCAAGCACCTGGGCGCTCAGTCAGTGGACCTCTGTTC 381
Db 101 eMetTyrIleSerGluThrAlaSerValHisLeuGlyLeuSerGln----- 116
QY 382 TCCTCCCTGTATACATACCCCACTCTCTGGTACCAATTTCTCTGGAGCTCATTTGACAC 441
;::: ||||| |||||
```

```
Db 117 -----Val--GluLeuThrGlyAsn 122
QY 442 AGTATCTTTTATCCATCTCTGTGACCAAGAGGAACCTTCAAGACGCCCTGACCCCC 501
Db 123 SerIleTyrGluTyrIleHisProSerAspHisAspGluMetThrAlaValLeuThrAla 142
QY 502 AGGCCGAACCTGTCAAGAAGAAAGCTG---GAAGCCCCAACAGAGCCCACTTTTCCCTG 558
Db 143 HisGlnProLeuHisHisLeuLeuGlnGluTyrGluIleGluArgSerPhePheLeu 162
QY 559 CGAATGAAGAGCAGCTCACCCAGAGGGCGGCACCTCAACCTCAAGAGCGGCCACCTGG 618
Db 163 ArgMetLysCysValLeuAlaLys-----ArgAsnAlaGlyLeuThrCysSerGlyTyr 180
QY 619 AAGTGTCTGCACCTCAGGACATATGAGGGCTACAAAGCCCCCTGCACAGACTCCCT 678
Db 181 LysValIleHisCysSerGlyTyrLeuLysIleArgGlnTyrMetLeuAspMetSerLeu 200
QY 679 GCCGGGAGC-----CCTCGCTCCGAGCCTCCCTG 708
Db 201 TyrAspSerCysTyrGlnIleValGlyLeuValAlaValGlyGlnSerLeuPro--- 219
QY 709 CAATGCTGTGCTTATCTGTGAAGCATCCCCAGCTCCCTTCCACGATGGT----- 762
Db 220 -----SerAlaIleThrGluIleLysLeuTyrSerAsnMetPhe 232
QY 763 -----GCTACTCTGGGTCTT----- 777
Db 233 MetPheArgAlaSerLeuAspLeuLysLeuIlePheLeuAspSerArgValThrGluVal 252
QY 778 -----CCACAG-----GAGAAGACTCCCATC-----TCT 801
Db 253 ThrGlyTyrGluProGlnAspLeuIleGluLysThrLeuTyrHisHisValHisGlyCys 272
QY 802 ACCTTATTCACCTCTTTGGGAAGGCACACTTGTCTGCTCAAGAGGTGGCTGTT--- 858
Db 273 AspValPheHisLeuArgTyrAlaHisHisLeuLeuValLysGlyGlnValThr 292
QY 859 -----CAGTGTCTACAGGGGAAAGG----- 879
Db 293 LysTyrTyrArgLeuLeuSerLysArgGlyTyrPValTrpValGlnSerTyrAlaThr 312
QY 880 -----ACTGAATCTCTCTCC-----TCATGGGTCTG 909
Db 313 ValValHisAsnSerArgSerArgProHisCysIleValSerValAsnTyrValLeu 332
QY 910 TGGGCCCTTAACCGGAAATTTGCTGCTAGGAGGAGTGAAGGACATGGCCACGCTA 969
Db 333 ThrGluIleGluTyrLysGlu-----LeuGlnLeu 342
QY 970 TCCTTACCCAGAAACCCCAAAATGCTCCAAACACCACCAATAAGACCTCTCTTGTGTAG 1029
Db 343 SerLeuGluGlnValSerThrAlaLysSerGlnAspSerTrpArgThrAlaLeuSerThr 362
QY 1030 GCACCAAGAAAAAATAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1083
Db 363 SerGlnGluThrArgLysLeuValLysProLysAsnThrLysMetLysThrLys 380
RESULT 9
US-09-925-297-863
; Sequence 863, Application US/09925297
; Patent No. US20020081659A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: PA105
; CURRENT APPLICATION NUMBER: US/09/925,297
; CURRENT FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCT/US00/05989
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/124,270
; PRIOR FILING DATE: 1999-03-12
; NUMBER OF SEQ ID NOS: 928
```





DB: 10 Gaps: 11

US-09-896-791b-2 (1-1100) x US-09-764-864-1139 (1-591)

QY 7 GGCACGAGGCCATGGCTGGTGGG-----CTGCAG 36  
||||| |||: |||  
Db 35 GGTThrArgProThrAlaMetGlySerPheSerHisMetThrGluPheProArgLys 54  
|||: |||: |||: |||  
QY 37 CGGTGAGGTTCGACACCGAGTCCGGAAGGAGGAGTCCGGGACGCGGCCCGCCAGCCGG 96  
|||: |||: |||: |||: |||  
Db 55 ArgLysGlySerAspSerAspProSerGlnGluAlaHisSerGlnThrGluLysArgArg 74  
|||: |||: |||: |||: |||  
QY 97 CGCAGCCAGAGACGAGGTGTGTACGAGTCCGCGACACTGCGCCCTTGGCGCGCGG 156  
|||: |||: |||: |||: |||  
Db 75 ArgAspLysMetAsnAsnLeuLeuGluLeuSerAlaMetileProGlnCysAsnPro 94  
|||: |||: |||: |||: |||  
QY 157 GTCAGCGCGCACCTGCAGCAGGCTCCATCATCGCTCACATACCTACCTACCTGCCCATG 216  
: |||: |||: |||: |||: |||: |||: |||: |||: |||: |||  
Db 95 MetAlaArgLysLeuAspLysLeuThrValLeuArgMetAlaValGlnHisLeuArg--- 113  
|||: |||: |||: |||: |||: |||: |||: |||: |||: |||  
QY 217 CACCGGCTCTGCGCAGCAGCTGGAAGGAGGAGGAGCCACTGGA----- 261  
: |||: |||: |||: |||: |||  
Db 114 -----SerLeuLysGlyLeuThrAsnSerTyrValGlySerAsnTyrArgPro 129  
|||: |||: |||: |||: |||: |||: |||: |||: |||: |||  
QY 262 -----CGCTGCTACCTGAAGCCCTGGAGGTTTCGT 294  
|||: |||: |||: |||: |||: |||: |||: |||: |||: |||  
Db 130 SerPheLeuGlnAspAsnGluLeuArgHisLeuLeuLeuLysThrAlaGluGlyPheLe 149  
|||: |||: |||: |||: |||: |||: |||: |||: |||: |||  
QY 295 CATGGTACTCACCGCCGAG---CGAGACATGGCTTACCTTCGGAANAATGTCACGAACA 351  
|||: |||: |||: |||: |||: |||: |||: |||: |||: |||  
Db 149 uPheValValGlyCysGluArgGlyLysLeuPheValSerLysSerValSerLysI 169  
|||: |||: |||: |||: |||: |||: |||: |||: |||: |||  
QY 352 CTTGGGCTCAGTCAGTGGACCTCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTG 411  
|||: |||: |||: |||: |||: |||: |||: |||: |||: |||  
Db 169 eLeuAsnTyrAspGln----- 174  
|||: |||: |||: |||: |||: |||: |||: |||: |||: |||  
QY 412 ACCAATTTCTCTGGAGCTGATGGACACAGATCTTTGATTTATTCATCCCTGTGAC 471  
|||: |||: |||: |||: |||: |||: |||: |||: |||: |||  
Db 175 -----Ala--SerLeuThrGlyGlnSerLeuPheAspPheLeuHisProLysAsp 190  
|||: |||: |||: |||: |||: |||: |||: |||: |||: |||  
QY 472 CAAGAGGAACCTCAAGACGCCCTG-----ACCCCGAGCCGAACCTG--- 513  
|||: |||: |||: |||: |||: |||: |||: |||: |||: |||  
Db 191 ValAlaLysValLysGluGlnLeuSerSerPheAspIleSerProArgGluLysLeuIle 210  
|||: |||: |||: |||: |||: |||: |||: |||: |||: |||  
QY 514 -----TCAAAGAAGAAGCTGGAA 531  
|||: |||: |||: |||: |||: |||: |||: |||: |||: |||  
Db 211 AspAlaLysThrGlyLeuGlnValHisSerAsnLeuHisAlaGlyArgThrArgValTyr 230  
|||: |||: |||: |||: |||: |||: |||: |||: |||: |||  
QY 532 GCCCAACAGAGCGCCATTTCCCTGCGAATGAAGACGACGCTCACCAGC----- 592  
|||: |||: |||: |||: |||: |||: |||: |||: |||: |||  
Db 231 SerGlySerArgArgSerPhePheCysArgIleLysSerCysLysIleSerValLysGlu 250  
|||: |||: |||: |||: |||: |||: |||: |||: |||: |||  
QY 583 ---AGAGGGCGCAGCTCAACCTCAAGCGCCACCTGGAG-----GTGCTGCAC 630  
|||: |||: |||: |||: |||: |||: |||: |||: |||: |||  
Db 251 GluHisGlyCysLeuProAsnSerLysLysLysGluHisArgLysPheTyrThrIleHis 270  
|||: |||: |||: |||: |||: |||: |||: |||: |||: |||  
QY 631 TGCTCAGGACATATAGGGCTTACAGCC-----CCTCAGCAGACTTCCCTGCGCGG 684  
|||: |||: |||: |||: |||: |||: |||: |||: |||: |||  
Db 271 CysThrGlyTyrLeuArgSerTyrProProAsnIleValGlyMetGluGluArgAsn 290  
|||: |||: |||: |||: |||: |||: |||: |||: |||: |||  
QY 685 AGCCCTCGCTCCGAGCCTCCCTCCCAATGCTGGTGTATCTGTGAAGCCATCCCCCGAG 744  
|||: |||: |||: |||: |||: |||: |||: |||: |||: |||  
Db 291 SerLysLysAspAsnSerAsnPheThrCysLeuValAlaIleGlyArgLeuGlnProTyr 310  
|||: |||: |||: |||: |||: |||: |||: |||: |||: |||  
QY 745 CTC---CCCTTCACATGGTGTCTCTCTGGGTCTTCCACAGGAG 786  
|||: |||: |||: |||: |||: |||: |||: |||: |||: |||  
Db 311 IleValProGlnAsnSerGlyGluIleAsnValLysProThrGlu 325  
|||: |||: |||: |||: |||: |||: |||: |||: |||: |||

RESULT 12  
US-09-864-761-41589  
: Sequence 41589, Application US/09864761  
: Patent No. US20020048763A1  
: GENERAL INFORMATION:

: APPLICANT: Penn, Sharron G.  
: APPLICANT: Rank, David R.  
: APPLICANT: Hanzel, David K.  
: APPLICANT: Chen, Wensheng  
: TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR  
: TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY  
: FILE REFERENCE: Aemica-X-1  
: CURRENT APPLICATION NUMBER: US/09/864,761  
: CURRENT FILING DATE: 2001-05-23  
: PRIOR APPLICATION NUMBER: US 60/180,312  
: PRIOR FILING DATE: 2000-02-04  
: PRIOR APPLICATION NUMBER: US 60/207,456  
: PRIOR FILING DATE: 2000-05-26  
: PRIOR APPLICATION NUMBER: US 09/632,366  
: PRIOR FILING DATE: 2000-08-03  
: PRIOR APPLICATION NUMBER: GB 24263.6  
: PRIOR FILING DATE: 2000-10-04  
: PRIOR APPLICATION NUMBER: US 60/236,359  
: PRIOR FILING DATE: 2000-09-27  
: PRIOR APPLICATION NUMBER: PCT/US01/00666  
: PRIOR FILING DATE: 2001-01-30  
: PRIOR APPLICATION NUMBER: PCT/US01/00667  
: PRIOR FILING DATE: 2001-01-30  
: PRIOR APPLICATION NUMBER: PCT/US01/00664  
: PRIOR FILING DATE: 2001-01-30  
: PRIOR APPLICATION NUMBER: PCT/US01/00669  
: PRIOR FILING DATE: 2001-01-30  
: PRIOR APPLICATION NUMBER: PCT/US01/00665  
: PRIOR FILING DATE: 2001-01-30  
: PRIOR APPLICATION NUMBER: PCT/US01/00668  
: PRIOR FILING DATE: 2001-01-30  
: PRIOR APPLICATION NUMBER: PCT/US01/00663  
: PRIOR FILING DATE: 2001-01-30  
: PRIOR APPLICATION NUMBER: PCT/US01/00662  
: PRIOR FILING DATE: 2001-01-30  
: PRIOR APPLICATION NUMBER: PCT/US01/00661  
: PRIOR FILING DATE: 2001-01-30  
: PRIOR APPLICATION NUMBER: PCT/US01/00670  
: PRIOR FILING DATE: 2001-01-30  
: PRIOR APPLICATION NUMBER: US 60/234,687  
: PRIOR FILING DATE: 2000-09-21  
: PRIOR APPLICATION NUMBER: US 09/608,408  
: PRIOR FILING DATE: 2000-06-30  
: PRIOR APPLICATION NUMBER: US 09/774,203  
: PRIOR FILING DATE: 2001-01-29  
: NUMBER OF SEQ ID NOS: 49117  
: SOFTWARE: Anomax Sequence Listing Engine vers. 1.1  
: SEQ ID NO 41589  
: TYPE: PRT  
: ORGANISM: Homo sapiens  
: FEATURE:  
: OTHER INFORMATION: MAP TO AL139296.2  
: OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 3  
: OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 2.9  
: OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 3.7  
: OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 2.7  
: OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 2.8  
: OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 3  
: OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 2.9  
: OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 2.9  
: OTHER INFORMATION: EST\_HUMAN HIT: A1360153.1, EVALUATE 1.00e-21  
: OTHER INFORMATION: SWISSPROT HIT: Q24119, EVALUATE 5.00e-21  
US-09-864-761-41589

Alignment Scores:  
Pred. No.: 8.73e-05 Length: 53  
Score: 153.00 Matches: 32  
Percent Similarity: 76.00% Conservative: 6  
Best Local Similarity: 64.00% Mismatches: 12  
Query Match: 7.61% Indels: 0  
DB: 10 Gaps: 0

US-09-896-791B-2 (1-1100) x US-09-864-761-41589 (1-53)

QY 58 CTGCGAAGGAGCACTCGCGGACCGCGCCCGCAGCGCGCAGGAGCGAGGTG 117  
Db 4 LeuArgLysGluLysSerArgAspAlaAlaArgSerArgGlyLysGluAsnPheGlu 23  
QY 118 CTGTACCACTGGCGCACACTGTGCCCTTTCGCGCGCGCGTCAAGGCGCACCTGGACAAG 177  
Db 24 PheTyrGluLeuAlaLysLeuLeuProLeuProAlaAlaIleThrSerClnLeuAspLys 43  
QY 178 GCCTCCATCATCGCGCTCACATCAGCTAC 207  
Db 44 AlaSerIleIleArgLeuThrIleSerTyr 53

## RESULT 13

US-09-800-729-112

; Sequence 112, Application US/09800729  
; Patent No. US20020068319A1  
; GENERAL INFORMATION:  
; APPLICANT: NI et al.  
; TITLE OF INVENTION: 32 Human secreted proteins  
; FILE REFERENCE: P2044P1  
; CURRENT APPLICATION NUMBER: US/09/800,729  
; CURRENT FILING DATE: 2001-03-08  
; PRIOR APPLICATION NUMBER: PCT/US00/26013  
; PRIOR FILING DATE: 2000-09-22  
; PRIOR APPLICATION NUMBER: 60/155,709  
; PRIOR FILING DATE: 1999-09-24  
; NUMBER OF SEQ ID NOS: 217  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 112  
; LENGTH: 530  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: SITE  
; LOCATION: (488)  
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
; NAME/KEY: SITE  
; LOCATION: (490)  
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
; NAME/KEY: SITE  
; LOCATION: (494)  
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
; NAME/KEY: SITE  
; LOCATION: (495)  
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
; NAME/KEY: SITE  
; LOCATION: (505)  
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
US-09-800-729-112

## Alignment Scores:

Prod. No.: 0.000672 Length: 530  
Score: 144.50 Matches: 121  
Percent Similarity: 29.67% Conservative: 41  
Best Local Similarity: 22.16% Mismatches: 121  
Every Match: 7.19% Indels: 264  
Gaps: 31

US-09-896-791B-2 (1-1100) x US-09-800-729-112 (1-530)

QY 111 GGAGTGTGTGA-----CCAGCTGGCGCACACTGTGCCCTTTCGCGCGGCGGT----- 158  
Db 27 GlyGlyAlaValGlnProGlySerLeuArgLeuSerCysAlaAlaSerGlyPhe 46  
QY 159 -----CAGCGCGCACCTGGACAAAGCGCTCCATCAT 188  
Db 47 ThrPheSerArgTyrGlyMethIleTrpValArgGlnAlaProGlyLysGlyLeuGln-Tr 66  
QY 189 G-----CG 191  
Db 66 pLeuAlaLeuValLeuHisAspGlyGlnLysTyrAsnGluAspValValLysGlyAr 86

QY 192 CCTCACAATCAGC-----TACCTGCGCATGCGACCGGCTCTGT 227  
Db 86 gPheThrIleSerArgAspAsnSerAsnAsnLysValTyrLeuGlnMetAspSerLeu-- 105  
QY 228 CGCAGAGGTGAAAGAGGGGAGAGCCACTGGAGCCCTGCTTACCTG----- 274  
Db 106 -----ArgGlyGluAspThrAla-ThrTyrTyrCysValArgGlyMet 120  
QY 275 --AAGGCCCTGGAGGTTTCGTCATGCTACTACCGCGGAGGAGACATGGCTTACCTGT 332  
Db 120 rpGluGlnLeuProSerTyrTyrPheAspTyrTrpGlyGlnGlyThrLeuValThrValS 140  
QY 333 CGGAAATGTACAGCAAGCAGCTGGCGCTCAGTCAGTGGACCTCTCTCCTCCTCCCTGAT 392  
Db 140 er-----SerAlaSerPro----- 144  
QY 393 ACATAACCCCACTCTCTGTACCAATTTCTCTGGAGCTCATTTGGACACAGTATCTTTGA 452  
Db 145 --ThrSerProLysValPheProLeuSerLeuCysSerThrGlnProAspGlyAsnVal 164  
QY 453 TTTTATCCATC----- 463  
Db 164 alIleAlaCysLeuValGlnGlyPhePheProGlnGluProLeuSerValThrTrpSerG 184  
QY 464 -----CCTGTGACCAAGAGGAACTTCAAGAGCGCCTGACCCCGAGCCGA--- 508  
Db 184 luSerGlyGlnGlyValThrAlaArgAsnPhe-----ProProSerGlnAspA 200  
QY 508 ----- 508  
Db 200 laSerGlyAspLeuTyrThrThrSerSerGlnLeuThrLeuProAlaThrGlnCysLeuA 220  
QY 509 -----ACCTGTCAAAGAAGCTGG---AAGCCCCAACAGAGCGCCACTT 551  
Db 220 laGlyLysSerValThrCysHisValLysHisTyrThrAsnProSerGlnAspValThrV 240  
QY 552 TTCCTCTGGAATGAAGAGCAGCCTCAGCAGAGAGGGGCGCAGCTCAACCTCAAAGCGCG 611  
Db 259 roSerProSerCysCys-----HisProArgLeuSerLeuHisArgp 273  
QY 612 CACCTGGAAGGTGCTGCATGCTCAGAGCATATGAGGCGCTACAAGCCCCCTGCACAGAC 671  
Db 259 roSerProSerCysCys-----HisProArgLeuSerLeuHisArgp 273  
QY 672 TTCCTCTGCGGGAGCCCTCGCTCGAGG-----CTCCCTGCAAT----- 712  
Db 273 roAlaLeuGluAspLeuLeuLeuGlySerGluAlaAsnLeuThrCysThrLeuThrGlyL 293  
QY 713 -----GCTGG----- 718  
Db 293 euArgAspAlaSerGlyValThrPheThrThrProSerSerGlyLysSerAlaValG 313  
QY 719 -----TGCTTATCTGTGAA-----GCATCCCGCCAGC 745  
Db 313 InGlyProProAspArgAspLeuCysGlyCysTyrSerValSerValLeuProGlyC 333  
QY 746 TC-----CCCTTCCACGATGGT-----GCTACTCTGGGTCTTCCACAGGAGA 787  
Db 333 ysAlaGluProTrpAsnHisGlyLysThrPheThrCysThrAlaAlaTyrProGluSerL 353  
QY 788 AGACTCCCATCTCTACC-----T 805  
Db 353 ysThrProLeuThrAlaThrLeuSerLysSerGlyAsnThrPheArgProGluValHisL 373  
QY 806 TATTCACCCCTCTTTGGAAG-----GCATCTACTTTGCTCTTG 841  
Db 373 euLeuProProProSerGluGluLeuAlaLeuAsnGluLeuValThrLeuThrCysLeuA 393  
QY 842 TCAAG-----AGTGGCCTTTCAGGTGCTTACAGGGA 874  
Db 393 laArgGlyPheSerProLysAspValLeuValArgTrp-----LeuGlnGly- 408





**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 Compugen Ltd.

OM nucleic - protein search, using frame\_plus\_n2p model

Run on: December 2, 2002, 20:32:35 : Search time 39 seconds  
(without alignments)  
5422.967 Million cell updates/sec

Title: US-09-896-791B-2  
Perfect score: 2010  
Sequence: 1 gaattcgccagcaggccat.....aaaaaaacatgcggccgc 1100

Scoring table: BLOSUM62  
Xgapop 10.0, Xgapext 0.5  
Ygapop 10.0, Ygapext 0.5  
Fgapop 6.0, Fgapext 7.0  
Delop 6.0, Delext 7.0  
Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 566448

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:  
-MODEL=frame\_n2p.model -DEV=xlp  
-O=/cgn2\_i/USPTO\_Spool/US09896791/runat\_02122002\_152658\_5606/app\_query.fasta\_1.1287  
-DB=PIR\_73 -QFMT=fastan -SUFFIX=rpr -MINMATCH=0.1 -LOOPEXT=0 -LOOPEXT=0  
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45  
-DOALIGN=200 -THR SCORE=pct -THR\_MAX=100 -THR\_MIN=0 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US09896791.ecgn\_1\_1\_30/runat\_02122002\_152658\_5606 -NCPU=6 -ICPU=3  
-NO\_XLPHY -NO\_MAP -LARGESOURCY -NEG\_SCORES=0 -WAIT -LONGLOG -DEV\_TIMEOUT=120  
-WARN\_TIMEOUT=30 -THRPADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7  
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : PIR\_73:\*  
1: pir1:\*  
2: pir2:\*  
3: pir3:\*  
4: pir4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	982.5	48.9	667	2 JC7771	hypoxia inducible
2	610.5	30.4	810	2 JC4837	hypoxia-inducible
3	607.5	30.2	811	2 JC7619	hypoxia-inducible
4	606.5	30.2	826	2 J38972	hypoxia-inducible
5	597.5	29.7	813	2 JC5809	hypoxia-inducible
6	349	17.4	1505	2 JC4851	hypoxia-inducible
7	306.5	15.2	248	2 A58520	single-minded gene
8	242.5	12.1	655	2 A29945	neurogenesis regul
9	216.5	10.8	823	2 T21943	hypothetical prote
10	216.5	10.8	825	2 T21944	hypothetical prote
11	172.5	8.6	300	2 T24292	hypothetical prote
12	166.5	8.3	392	2 JC7633	aryl hydrocarbon n
13	162	8.1	805	2 JC7635	aryl hydrocarbon r
14	158.5	7.9	791	2 A56241	aryl hydrocarbon r

## ALIGNMENTS

RESULT 1  
JC7771  
hypoxia inducible factor-3 alpha - human  
C:Species: Homo sapiens (man)  
C>Date: 01-Feb-2002 #sequence\_revision 01-Feb-2002 #text\_change 01-Feb-2002  
C:Accession: JC7771  
R:Hara, S.; Hamada, J.; Kobayashi, C.; Kondo, Y.; Imura, N.  
Biochem. Biophys. Res. Commun. 287, 808-813, 2001  
A:Title: Expression and characterization of hypoxia-inducible factor (HIF)-3alpha in  
A:Reference number: JC7771; PMID:11573933  
A:Contents: Kidney  
A:Accession: JC7771  
A:Molecule type: mRNA  
A:Residues: 1-667 <HAR>  
C:Cross-references: DDBJ:AB054067  
C:Comment: This protein is a heterodimeric transcription factor that belongs to the b  
lved in the regulation of hypoxia-inducible gene expression in human kidney.  
C:Genetics:  
A:Gene: hif-3alpha  
A:Map position: 19  
C:Keywords: kidney

Alignment Scores:	Pred. No.:	Length:
Score:	1.38e-70	667
Percent Similarity:	982.50	205
Best Local Similarity:	85.08%	Conservative: 6
Query Match:	82.66%	Mismatches: 7
DB:	48.88%	Indels: 30
	2	Gaps: 3
US-09-896-791B-2 (1-1100) x JC7771 (1-667)		
QY 43	AGGTGCAACACCGAGCTGCGGAGGAGAGTGCAGGACCGCGCGCGCGCGCGCGCAGC	102
Db 7	ArgSerThrThrGluLeuArgLysGluTyrSerArgAspAlaAlaArgSerArgSer	26
QY 103	CAGGAGACGGAGGTGCTGTACCGAGTGGCGGCACACTCTCCCTTTTTCGCGCGCGCGCGCAGC	162

brain and muscle A  
Ah receptor nuclea  
aryl hydrocarbon r  
Arnt-like PAS prot  
aryl hydrocarbon r  
aryl hydrocarbon r  
aryl hydrocarbon r  
aryl hydrocarbon r  
aryl hydrocarbon r  
aryl hydrocarbon r  
aryl hydrocarbon r  
cyclin T - fruit f  
protein C4IG7.5 [1  
hypothetical prote  
brain and muscle A  
tegument protein 6  
neurogenesis regul  
glucocorticoid rec  
probable large pro  
brain and muscle A  
proline-rich protei  
probable proline-r  
unconventional myo  
hypothetical prote  
cell proliferation  
hypothetical 70K p  
ORF2 protein - Orf  
eyelid - fruit fly

```
Db 27 GlnGluThrGluValLeuTyArgInLeuAlaHisThrLeuProPheAlaArgGlyValSer 46
QY 163 GCGCACCTGGCAAGCGCTCATCATCGCGCTCACAAATCAGCTACCTGCGCATGCGCACCGC 222
Db 47 AlaHisLeuAspLysAlaSerIleMetArgLeuThrIleSerTyArgMetHisArg 66
QY 223 CTCTGCGCAGCAGGT-----GGAAAGGGGGAGAGCCACTGGACGCCCTG 267
Db 67 LeuCysAlaAlaGlyGluTrpAsnGlnValGly-AlaGlyGlyGluProLeuAspAlaCys 86
QY 268 CTACCTGAAGGCCCTGGAGGTTTCGTACGTACTCAGCCCGGAGGAGACATGCTTGA 327
Db 86 sTyLeuLysAlaLeuGluGlyPheValMetValLeuThrAlaGluGlyAspMetAlaTy 106
QY 328 CTGTGCGGAAATGTGAGCAAGCACCTGGGCCCTCAGTCAGTGGACCTCTGTTCTCCTCCTCC 387
Db 106 rLeuSerGluAsnValSerLysHisLeuGlyLeuSerGln----- 119
QY 388 CTGATACATAACCCCACTCCTGCTACCAATTTCTCTCTGGAGCTCATTTGGACACAGTATC 447
Db 120 ----- 127
QY 448 TTTGATTTTATCCATCCCTGTGACCAAGAGAACTTCAAGACGCCCTGACCCCGGCGG 507
Db 128 PheAspPheIleHisProCysAspGlnGluLeuGlnAspAlaLeuThrProGlnGln 147
QY 508 AACCTGTCAAAGAAAGAGCTGGAAGCCCAACAGAGCGCCACTTTTCCCTGCCAATGAAG 567
Db 148 ThrLeuSerArgGlyValGluAlaProThrGluArgCysPheSerLeuArgMetLys 167
QY 568 AGCAGCGTCACACGACAGAGGGCGCAGCTCAACCTCAAGCGGCCACCTGGAAAGGTGCTG 627
Db 168 SerThrLeuThrSerArgGlyArgThrLeuAsnLeuLysAlaAlaThrTrpLysValLeu 187
QY 628 CACTGCTCAGGACATATAGGGCTACAGCCCGCTCAACCTCAAGCGGCCACCTGGAAAGGTGCTG 687
Db 188 AsnCysSerGlyHisMetArgAlaTyLysProProAlaGlnThrSerProAlaGlySer 207
QY 688 CFTCGCTCCGAGCGCTCCCTGCAATGCTGTGCTTATCTGTGAAGCCATCCCGCCAGCTC 747
Db 208 ProAspSerGluProLeuGlnCysLeuValIleCysGluAlaIlePro----- 225
QY 748 CCCTTCACGATGGTCT 765
Db 226 -----HisProGlySer 229
RESULT 2
JC4837
hypoxia-inducible factor 1 alpha - mouse
C:Species: Mus musculus (house mouse)
C:Date: 15-Aug-1996 #sequence_revision 15-Oct-1996 #text_change 21-Jul-2000
C:Accession: JC4837
R:Wenger, R.H.; Rolfs, A.; Marti, H.H.; Guenet, J.L.; Gassmann, M.
Biochem. Biophys. Res. Commun. 223, 54-59, 1996
A:Title: Nucleotide sequence, chromosomal assignment and mRNA expression of mouse hypoxi
A:Reference number: JC4837; MUID:96254028; PMID:8660378
A:Accession: JC4837
A:Molecule type: mRNA
A:Residues: 1-810 <WEN>
A:Cross-references: EMBL:X95580; NID:g1430864; PIDN:CAA64833.1; PID:g4379202
A:Comment: This factor is involved in the oxygen-regulated transcription of several gene
C:Genetics:
A:Gene: Hif1alpha
A:Map position: 12
C:Keywords: transcription factor
F:5-58/Region: helix-loop-helix #status predicted
Alignment Scores:
Pred. No.: 7.84e-41 Length: 810
Score: 610.50 Matches: 129
Percent Similarity: 67.65% Conservative: 32
Best Local Similarity: 54.20% Mismatches: 49
Query Match: 30.37% Indels: 28
```

```
DB: 2 Gaps: 3
US-09-896-791b-2 (1-1100) x JC4837 (1-810)
QY 49 AACACCGAGCTCGGGAAGAGAGAGTCCGGGAGCGGCCGCGCAGCGCGCAGCAGGAG 108
Db 2 SerSerGluArgArgLysGluLysSerArgAspAlaAlaArgSerArgSerLysGlu 21
QY 109 ACGGAGGTCTGTACACAGCTGGCGCACACTCTGCTCCCTTTGCGGCGCGCTGCACGCGCAC 168
Db 22 SerGluValPheTyArgGluLeuAlaHisGlnLeuProLeuProHisAsnValSerSerHis 41
QY 169 CTGGACAAGCCCTCCATCATCGCCCTCACAAATCAGTACTGCGCATGCGCACCGCCCTCTCC 228
Db 42 LeuAspLysAlaSerValMetArgLeuThrIleSerTyArgValArgLysLeuLeu 61
QY 229 GCAGCAGGTGGAAAAGGGGGAGAG-----CCACTGGAGCCCTGCTACCT 273
Db 62 AspAlaGlyGly-LeuAspSerGluAspGluMetLysAlaGlnMetAspCysPheTyLe 81
QY 274 GAAGGCCCTGGAGGTTTCGTATGCTACTCACCGCGGAGGAGACATGGCTTACCTGTC 333
Db 81 uLysAlaLeuAspGlyPheValMetValLeuThrAspAspGlyAspMetValTyIleSe 101
QY 334 GGAAATGTGACAGCACCTGGCCCTCAGTCAGTCAGTCAGTCTGTTCTCTCCCTCCCTGATA 393
Db 101 rAspAsnValAsnLysTyArgMetGlyLeuThrGln----- 112
QY 394 CATAACCCCACTCCTGTTACCAATTTCTCTGAGACTCATTTGGACACATATCTTTGAT 453
Db 113 -----Phe--GluLeuAlaGlyHisSerValPheAsp 122
QY 454 TTTATCCATCCCTGTGACCAAGAGAACTTCAAGACGCCCTGACCCCGAGCGGAACCTG 513
Db 123 PheThrHisProCysAspHisGluGluMetArgGluMetLeuThrHisArgAsnGlyPro 142
QY 514 TCAAAGAAGAAGCTGGAAGCCCCAACAGAGCGCCACTTTTCCCTGCGAATGAAGACACG 573
Db 143 ValArgLysGlyLysGluLeuAsnThrGlnArgSerPhePheLeuArgMetLysCysThr 162
QY 574 CTCACACGAGGCGGCGCAGCTCAACCTCAAGCGCGCCACCTGGAAGTGTGCTGCTGCTG 633
Db 163 LeuThrSerArgGlyArgThrMetAsnIleLysSerAlaThrTrpLysValLeuHisCys 182
QY 634 TCAGGACATATGAGGCGCTACAAAGCCCGCTGCACAGACTTCCCTCGCGGAGCCCTCGC 693
Db 183 ThrGlyHisIleHisValTyArgThrAsnSerAsnGlnProGlnCysGly-----Tyr 200
QY 694 TCGAGCCCTCCCTGCAATGCGCTGTGCTTATCTGTGAAGCCCATCCCC 741
Db 201 LysLysProProMetThrCysLeuValLeuIleCysGluProIlePro 216
RESULT 3
JC7619
hypoxia-inducible factor 1 alpha - chicken
C:Species: Gallus gallus (chicken)
C:Date: 30-Jun-2001 #sequence_revision 30-Jun-2001 #text_change 30-Jun-2001
C:Accession: JC7619
R:Takahashi, T.; Sugishita, Y.; Nojiri, T.; Shimizu, T.; Yao, A.; Kinugawa, K.; Harad
Biochem. Biophys. Res. Commun. 281, 1057-1062, 2001
A:Title: Cloning of hypoxia-inducible factor 1 alpha cDNA from chick embryonic ventri
A:Reference number: JC7619; MUID:21134360; PMID:1123772
A:Contents: Embryonic ventricular myocytes
A:Accession: JC7619
A:Molecule type: mRNA
A:Residues: 1-811 <TAK>
A:Cross-references: DDBJ:AB013746
A:Comment: This factor belongs to the basic helix-loop-helix-Per/ARNT/Sim (HLH-PAS) f
eostasis of cells, and in redox stimuli.
C:Genetics:
A:Gene: hif-1alpha
C:Keywords: embryo; transcription factor
F:106-156/Domain: Per/ARNT/Sim, ligand binding, dimerization #status predicted <PAS>
```



F; 249-299/Domain: Per/ARNT/Sim, ligand binding, dimerization #status predicted <PAS2>  
F; 762-811/Domain: conserved carboxy-terminal transactivation element #status predicted <PAS2>  
F; 767-768/Region: conserved dileucine repeat, important for oxygen-dependent degradation

Alignment Scores:		
Pred. No.:	1,36e-40	811
Score:	607.50	136
Percent Similarity:	67.36%	25
Best Local Similarity:	56.90%	Conservative: 51
Query Match:	30.22%	Mismatches: 28
DB:	2	Indels: 3
		Gaps: 3

US-09-896-791B-2 (1-1100) x JC7619 (1-811)

QY	43	AGTGTCAACACCAGCTGCGGAGGAGAGTAGTCGCGGGACGCGGCCCGCAGCCGGCGCAGC	102
		: : : :	
Db	12	ArgIleSerSerGluArgArgLysGluLysSerArgAspAlaAlaArgCysArgArgSer	31
QY	103	CAGGACGAGGAGTGTGTACCAAGCTGGCGGCACACTCTCCCTTTTCGCGCGCGCTCAGC	162
		: : : : :	
Db	32	LysGluSerGluValPheTyrGluLeuAlaHisGluLeuProLeuProHisThrValSer	51
QY	163	GCCCACTGTGACAAAGGCCTCCATCATCGCGCTCACAAATCAGCTACCTGCGCATGCACCGC	222
Db	52	AlaHisLeuAspLysAlaSerIleMetArgLeuThrIleSerTyrLeuArgMetArgLys	71
QY	223	CTCTGCGCAGCAGGTGGAAAGGGGGAGAG-----CCACTGGACGCGCTG	267

72	LeuLeuAspAlaGly-GluLeuGluThrGluAlaAsnMetGluLeuGluLeuAsnCys	91
268	CTACCTGAGGCCCTCGAGGTTTCCTCATGGTACTACCGCGGAGGAGACATGGCTTA	327
91	eYrLeuLysAlaLeuAspGlyValMetValLeuSerGluAspGlyAspMetIleTy	111
328	CCCTGCGAAATGTGACGAACACCTCGGCCCTCACTAGTG-GACCTCTGTCTCCCTCC	386
111	rMetSerGluAsnValAsnLysCysMetGlyLeuThrGlnPheAsp-----	126

Qy	387	CCTGATACATAACCCACCTCTCTGGAGTCAATGGACACAGTAT	446
		:::	
Db	127	-----LeuThrGlyHisSerVa	132

Qy 447 CTTTGAATTTATCCATCCCTGTGACCAAGAGGAACCTTCAAGACGCCCTGACCCCGAGGCC 506  
:||||| ||||||| |||||||:|:| ||||| |||  
Db 132 lPheAspPheThrHisProCysAspHisGluGluLeuArgGluMetLeuThrHisArgAs 152

QY 507 GAACCTGTCAAAGAAGAGCTGGAAGCCCCAACAGAGCGCCACTTTTCCCTGCGAATGAA 566  
||||| ||| ||||||| ||| |||||||  
Db 152 nGlyProValLysLysGlyLysGluclnAsnThrGluAraSerPhePheLeuAraMetIy 172

QY 567 GAGCAGCGTCACCAGCAGGAGGGCGGACGCTCAACCTCAAAGCGGCCACTGGAAGGTGCT 626  
- | ||||| ||||||| ||||||| :|||:::|||:||||| |||||  
Db 172 sCvstThrLeuthSerArgGLVArGThrValAsnILtElvsSerAlathTrpLysValle 192

Qy\* 627 GCACTGCTCAGGACATATGAGGGCCCTACAGCCCCCTGCACAGACTTCCCCCTGCGGGGAG 686  
|||||:::|||||:::||||| ||| ::|  
Db 192 uHisCvsthrG|vh|s|I|AraVa|TvrAspThrCvsaAsnAspC|nThrHisCvseGlv-- 211  
|||||:::|||||:::||||| ||| ::|

687 CCTCCGCTCCGAGCCCTCCCTGCCAATGCCCTGCTGCTTATCTCTCTCAAGCCATACCCC 741

212 ----TyrIvslvsproProwMetThrCvslEnVa|LeuIlleCvsl|nproI|epr 228

DECIIT A

I38972

N: Alternate names: ARNT interacting proteins

C;Species: Homo sapiens (man)

C;Date: 23-Feb-1996 #sequence\_revision 23-F-20020 001075

C;Accession: 1389/2; GU18/5  
B:Wang, G.L.: Jiang, B.H.: Rue, E.A.: Semenza, G.L.

Proc. Natl. Acad. Sci. U.S.A.

**A;Title:** Hypoxia-inducible factor 1 is a basic-heli

A;Reference number: I38972; MUID:95296340; PMID:7

A:Accession: I38972  
A:Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-826 <RES>  
A:Cross-references: EMBL:U22431; NID:g881345; PID:AA05152.1; PID:g881346  
A:Note: parts of this sequence were confirmed by peptide sequencing  
R:Hognessch, J.B.; Chan, W.K.; Carver, L.A.; Bradford, C.A.  
submitted to the EMBL Data Library, June 1995  
A:Reference number: H00692

A;Accession: G01675  
A;Status: preliminary; translated from GB/EMBL/DDBJ

A; MOLECULE TYPE: IIIKNA  
A: Residues: 1-826 <HOG>

A;cross-references: EMBL:029105; NID:91144012; PDB:1AC31Z10.1; PDB:91144010.1	
C;Genetics:	
A;Gene: GDB:H1FLA	
A;Cross-references: GDB:512229	
A;Map position: 14q21-14q24	
C;Keywords: heterodimer	
Alignment Scores:	
Pred. No.:	1.64e-40
Score:	606.50
Percent Similarity:	68.07%
Best Local Similarity:	54.20%
Query Match:	30.17%
DB:	2
Gaps:	3
Length:	826
Matches:	129
Conservative:	33
Mismatches:	48
Indels:	28
Gaps:	3

US-09-896-791B-2 (1-1100) x I38972 (1-826)

Ov 49 AACACCGAGCTCGGGAGGAGAGTCTCGGGGAGCGGGCCGACCGGGCGCAGCCAGGAG 108

Db 14 SerSerGluArgLysGluLysSerArgAspAlaAlaArgSerArgArgSerLysGlu 33  
Ov 109 ACGGAGTGTCTATCCAGCTGGCGGCACACTCTGTCCTTTGCGCGCGGCTCAGGCGGCAC 168

Db 34 SerGluValPheTyrGluLeuAlaHisGlnLeuProLeuProHisAsnValSerSerHis 53  
Ov 169 CTGGGACAAGGCCTCCATCATGTGGCCCTACAATCAGCTACCTGGCATGACCCGCTCTGC 228

Db 54 LeuAspLysAlaSerValMetArgLeuThrIleSerTyrLeuArgValArgLysLeuLeu 73  
 QV 229 GCAGCAGCTGGAAAAAGSGGGGAG-----CCACTGGACGCTGCTACCT 273





```
Db 163 ArgMetLysCysValLeuAlaLys-----ArgAsnAlaGlyLeuThrCysSerGlyTyr 180
QY 619 AAGGTGCTGCACCTGCTCAGGACATATGAGG 648
Db 181 LysValIleHisCysSerGlyTyrLeuLys 190
RESULT 8
A29945
neurogenesis regulatory protein - fruit fly (Drosophila melanogaster) (fragment)
N:Alternate names: single-minded gene protein
C:Species: Drosophila melanogaster
C:Date: 15-Dec-1988 #sequence_revision 15-Dec-1988 #text_change 20-Mar-1998
C:Accession: A29945
R:Cell 52, 143-151, 1988
R:Crews, S.T.; Thomas, J.B.; Goodman, C.S.
A:Title: The Drosophila single-minded gene encodes a nuclear protein with sequence simil
A:Reference number: A29945; MUID:88151023; PMID:3345560
A:Accession: A29945
A:Molecule type: mRNA
A:Residues: 1-655 <CRE>
A:Cross-references: GB:M19020; NID:gl58464; PID:gl58465
C:Genetics:
A:Gene: sim
A:Cross-references: FlyBase:FBgn0004666
C:Keywords: DNA binding; transcription regulation
Alignment Scores:
Pred. No.: 2,11e-11 Length: 655
Score: 242.50 Matches: 76
Percent Similarity: 45.56% Conservative: 42
Best Local Similarity: 29.34% Mismatches: 67
Query Match: 12.06% Indels: 74
DB: 2 Gaps: 7
US-09-896-791b-2 (1-1100) x A29945 (1-655)
QY 124 CAGCTGGCGCACACTGTGCTTTCGGCGCGGCTCAGCGCGACCTGGACAGGCGCTCC 183
Db 4 GluLeuAlaLysLeuLeuProAlaIleThrSerGlnLeuAspLysAlaSer 23
QY 184 ATCATGGCGCTCACATACAGTACCTGCGCGATGCGCGCTGCGCAGCAGGT----- 237
Db 24 ValIleArgLeuThrThrSerTyrLeuLysMetArgGlnValPheProAspGlyLeuGly 43
QY 238 -----GGAAAAAGGGGAGGAGCCACTGGACCC 264
Db 44 GluAlaTrpGlySerSerProAlaMetGlnArgGlyAlaThrIleLysGluLeuGlySer 63
QY 265 CTGCTACCTGAAGGCCCTGGAGGGTTTCGTCTGCTACTCAACCGCGAGGAGACATGCG 324
Db 64 HisLeu-LeuGlnThrLeuAspGlyPheIlePheValValAlaProAspGlyLysIleMe 83
QY 325 TTACCTGTCGGAATGTTCAGACAGCACCTGGCGCTCAGTCACTAGTGGACCTCTGTCCTCC 384
Db 83 tTyrIleSerGluThrAlaSerValHisLeuGlyLeuSerGln----- 97
QY 385 TCCTGATACATAACCCCACTCTGTGTACCAATTTCTCTCTGAGCTCATTGACACAGT 444
Db 98 -----Val--GluLeuThrGlyAsnSer 104
QY 445 ATCTTTGATTTTATCCATCCCTGTGACCAAGAGAACTTCAAGAGCGCCCTGACCCCCAGG 504
Db 105 IlePheGluTyrIleHisAsnTyrAspGlnAspGluMetAsnAlaIleLeuSerLeuHis 124
QY 505 CGGAACCTGTCAAAGAGAGCTG----- 528
Db 125 ProHisIleAsnGlnHisProLeuAlaGlnThrHisThrProIleGlySerProAsnGly 144
QY 529 ---GAAGCCCAACA-----GAGGCC 546
Db 145 ValGlnHisProSerAlaTyrAspHisAspArgGlySerHisThrIleGluLeuGlyLys 164
QY 547 CACTTTTCCCTCGGAATGAAGACGACGCTCACACAGAGGCGCAGCGCTCAACCTCAA 606
```

```
Db 165 ThrPhePheLeuArgMetLysCysValLeuAlaLys-----ArgAsnAlaGlyLeuThr 182
QY 607 GCGGCGCACCTGGAAGGTGCTGCACATATGAGGCGCTTACAGGCCCTGCA 666
Db 183 ThrSerGlyPheLysValIleHisCysSerGlyTyrLeuLysAlaArgIleTyrProAsp 202
QY 667 CAGACTTCCCTGCGGGAGCGCTCGCTCCGAGCGCTCCCAATGCTGCTGCTTATC 726
Db 203 ArgGlyAspGlyGlnGlySer-----LeuIleGlnAsnLeuGlyLeuVal 217
QY 727 TGT-----GAAGCCATCCCGCCAGCTCCCTCCAC 756
Db 218 AlaValGlyHisSerLeuProSerSerAlaIleThrGluIleLysLeuHis 234
RESULT 9
T21943
hypothetical protein F38A6.3a - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 29-Oct-1999
C:Accession: T21943; T26899
R:Mortimore, B.
submitted to the EMBL Data Library, March 1997
A:Reference number: Z19491
A:Accession: T21943
A:Status: preliminary; translated from GB/EMBL/DBDJ
A:Molecule type: DNA
A:Residues: 1-823 <WIL>
A:Cross-references: EMBL:Z92833; PIDN:CAB07380.1; GSPDB:GN00023; CESP:F38A6.3a
A:Experimental source: clone F38A6
R:Ainscough, R.
submitted to the EMBL Data Library, June 1998
A:Reference number: Z20282
A:Accession: T26899
A:Status: preliminary; translated from GB/EMBL/DBDJ
A:Molecule type: DNA
A:Residues: 1-823 <WI2>
A:Cross-references: EMBL:AL023842; PIDN:CAA19520.1; GSPDB:GN00023; CESP:F38A6.3a
A:Experimental source: clone Y44A6D
C:Genetics:
A:Gene: CESP:F38A6.3a
A:Map position: 5
A:Introns: 19/2; 75/1; 111/2; 175/1; 226/3; 331/3; 518/1; 627/2; 711/1; 768/3
Alignment Scores:
Pred. No.: 2,55e-09 Length: 823
Score: 216.50 Matches: 69
Percent Similarity: 47.33% Conservative: 46
Best Local Similarity: 28.40% Mismatches: 66
Query Match: 10.77% Indels: 63
DB: 2 Gaps: 9
US-09-896-791b-2 (1-1100) x T21943 (1-823)
QY 43 AGGTGCAACACGAGTGTGGAAGGAGAGTGTGCGGAGCGCGCGCGAGCGGCGCAGC 102
Db 110 LysArgAsnMetGluArgArgGluThrSerArgHisAlaAlaArgAspArgSer 129
QY 103 CAGGAGACGGAGGTGCTGTACAGTGTGGGCGACACTCTGCGCTTTCGCG---CGCGGCGTC 159
Db 130 LysGluSerAspIlePheAspAspLeuLysMetCysValProIleValGluGluGlyThr 149
QY 160 AGCGCGCACCTGGCAAGCGCTCCATCATGCGCTCACATCAGTACCTACCTGCGCATGCAC 219
Db 150 ValThrHisLeuAspArgIleAlaLeuLeuArgValAlaAlaThrIleCysArgLeuArg 169
QY 220 CGCTCTCTCGCAGCAGGTGTGAAAAAGGGGGA----- 250
Db 170 Lys-ThrAlaGlyAsnValLeuGluAsnAsnLeuAspAsnGluIleThrAsnGluValTr 189
QY 251 -----GAGCCACTGGAGCGCTGTACTGGAAGCGCTTGGAGGGTTCGTGCTCATGCTACT 303
Db 189 pThrGluAspThrIleAlaGluCys-----LeuAspGlyPheValMetIleVal 205
```

Qy	304	CACGCCGAGGGAGACATGCCTTACTGTCTGGAAAAATGTCAGCAAGCACCTGGCGCCTCAG	363
:	:	: : : : : :	:
Db	205	IAspSerAspSerSeriLeuThrValThrGluSerValAlaMetTyrLeuGlyLeuTh	225
Qy	364	TCAGTG-GACCTCTGTTCCCTCCTGATACATAACCCCACTCCTGGTACCATTTC	422
:	:	: : : : : :	:
Db	225	rGlnThrAsp-	228
Qy	423	TCTGGAGCTCATTTGGACACAGTATCTTTGATTCTTATCCATCCCTGTGCACAAAGAGA	482
:	:	: : : : : :	:
Db	229	- - - - - LeuThrGlyArgAlaLeuArgAspPheLeuHisProSerAspTyrAspGluPh	246
Qy	483	TCAAGAGCCCTCACCACCCAGCCGAACCTGTCAAAGAAGAGCTGGAGAGCCCCA- - - -	537
:	:	: : : : : :	:
Db	246	e- - - - - AspLysGlnSerLysMetLeuHisLysProArgGl	258
Qy	538	- - - - - ACAGAGCGCCACTTTCCCTCGCAATGAAGAGCACACCTCACCAG	581
:	:	: : : : : :	:
Db	258	yGluAspThrAspThrThrGlyIleasnMetValLeuArgMetLysThrValIleSerPr	278
Qy	582	CAGAGCGCGCACGCTCAACCTCAAGCGCCACCTGGAAG- - - - - GTGCTGCA	629
:	:	: : : : : :	:
Db	278	oAtgGlyArgCysLeuAsnLeuLysSerAlaLeuTyrLysSerValSerPheLeuValHi	298
Qy	630	C- - - - - TGCTCAGGACATATGAGGGCTACAAGCCCCTGACAGACTTCCCC	677
:	:	: : : : : :	:
Db	298	sSerLysValSerThrGlyGlyHisValSerPheMetGln- - - - GlyIleThrIlePr	316
Qy	678	TGCGCGG	684
:	:	: : : : : :	:
Db	316	oAlaGly	318
 RESULT 10			
T21944			
hypothetical protein F38A6.3b - Caenorhabditis elegans			
C:Species:Caenorhabditis elegans			
C>Date:15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 29-Oct-1999			
C:Accession:T21944; T26900			
R:Mortimore, B.			
submitted to the EMBL Data Library, March 1997			
A:Reference number: Z19491			
A:Accession: T21944			
A>Status: preliminary; translated from GB/EMBL/DDBJ			
A:Molecule type: DNA			
A:Residues: 1-825 <WIL>			
A:Cross-references: EMBL:Z92833; PIDN:CAB07381.1; GSPDB:GN00023; CESP:F38A6.3b			
A:Xperimental source: clone F38A6			
R:Ainscough, R.			
submitted to the EMBL Data Library, June 1998			
A:Reference number: Z20282			
A:Accession: T26900			
A>Status: preliminary; translated from GB/EMBL/DDBJ			
A:Molecule type: DNA			
A:Residues: 1-825 <WI2>			
A:Cross-references: EMBL:AL023842; PIDN:CAA19521.1; GSPDB:GN00023; CESP:F38A6.3b			
A:Xperimental source: clone Y44A6D			
C:Genetics:			
A:Gene: CESP:F38A6.3b			
A:Map position: 5			
A:Introns: 19/2; 75/1; 111/2; 175/1; 226/3; 331/3; 518/1; 627/2; 713/1; 770/3			

Alignment Scores:		
Pred. No.:	2,55e-09	Length:
Score:	216.50	Matches:
Percent Similarity:	47.33%	Conservative:
Best Local Similarity:	28.40%	Mismatches:
Query Match:	10.77%	Indels:
DB:	2	Gaps:
		9
		825

US-09-896-791B-2 (1-1100) x T21944 (1-825)

QY 43 AGGTCGAACACCGAGCTGCGGAAGGAGAAAGTCGGGGACGGGGCCCGCAGCCGGCGCAGC 102

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	100	101	102	103	104	105	106	107	108	109	110	111	112	113	114	115	116	117	118	119	120	121	122	123	124	125	126	127	128	129	130	131	132	133	134	135	136	137	138	139	140	141	142	143	144	145	146	147	148	149	150	151	152	153	154	155	156	157	158	159	160	161	162	163	164	165	166	167	168	169	170	171	172	173	174	175	176	177	178	179	180	181	182	183	184	185	186	187	188	189	190	191	192	193	194	195	196	197	198	199	200	201	202	203	204	205	206	207	208	209	210	211	212	213	214	215	216	217	218	219	220	221	222	223	224	225	226	227	228	229	230	231	232	233	234	235	236	237	238	239	240	241	242	243	244	245	246	247	248	249	250	251	252	253	254	255	256	257	258	259	260	261	262	263	264	265	266	267	268	269	270	271	272	273	274	275	276	277	278	279	280	281	282	283	284	285	286	287	288	289	290	291	292	293	294	295	296	297	298	299	300	301	302	303	304	305	306	307	308	309	310	311	312	313	314	315	316	317	318	319	320	321	322	323	324	325	326	327	328	329	330	331	332	333	334	335	336	337	338	339	340	341	342	343	344	345	346	347	348	349	350	351	352	353	354	355	356	357	358	359	360	361	362	363	364	365	366	367	368	369	370	371	372	373	374	375	376	377	378	379	380	381	382	383	384	385	386	387	388	389	390	391	392	393	394	395	396	397	398	399	400	401	402	403	404	405	406	407	408	409	410	411	412	413	414	415	416	417	418	419	420	421	422	423	424	425	426	427	428	429	430	431	432	433	434	435	436	437	438	439	440	441	442	443	444	445	446	447	448	449	450	451	452	453	454	455	456	457	458	459	460	461	462	463	464	465	466	467	468	469	470	471	472	473	474	475	476	477	478	479	480	481	482	483	484	485	486	487	488	489	490	491	492	493	494	495	496	497	498	499	500	501	502	503	504	505	506	507	508	509	510	511	512	513	514	515	516	517	518	519	520	521	522	523	52
--	---	---	---	---	---	---	---	---	---	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	----

Score: 172.50 Matches: 59  
 Percent Similarity: 46.82% Conservative: 44  
 Best Local Similarity: 26.82% Mismatches: 70  
 Query Match: 8.58% Indels: 47  
 DB: 2 Gaps: 6

US-09-896-791B-2 (1-1100) x T24292 (1-300)

QY 124 CAGCTGGCAGCACTCTGGCGGCGGTCTAGCGCG---CACCTGGACAAGGCC 180  
 Db 8 GlnLeuAlaAsnGlnLeuProLeuAlaAlaIleSerGlyGlnHisIleAspLysThr 27  
 QY 181 TCCATCATGCGCCTCACAAATCAGCTACCTCGCATGCGACCGCTCTGCGAGCAGGTGGA 240  
 Db 28 ThrMetValArgLeuAlaThrAlaTyIleLysLeuHisAsnIlePheGly-GlnSerGI 47  
 QY 241 AAAAGGGGAGACCACTGACGCGCTCTAC----- 271  
 Db 47 nArgAlaTySerSerAlaAspTyTyTyGlySerAspSerLeuTrpThrAsnAsnHi 67  
 QY 272 -CTGAAGCGCCTGGAGGCTTCGTCTACTCAGCGCGGAGGAGACATGGCTTACCT 330  
 Db 67 sLeuAspLeuLeuAspGlyPhePheValIleLeuAspArgArgGlyAspValLeuTyriI 87  
 QY 331 GTCGGAAATGTGACGAAGCACTGCGGCTGACGTGAGTGGACCTCTCTCTCTCTCTCT 390  
 Db 87 eSerGluThrIleSerIleTyLeuGlyLeuSerGln----- 99  
 QY 391 ATACATAACCCCACTCTCTGTACCAATTTCTCTGTGAGCTCATTTGACACAGATATCTT 450  
 Db 100 -----Val--GluMetThrGlyAsnAlaMetVal 108  
 QY 451 GATTTTATCCATCCTGTGACCAAGGAACCTTCAAGACGCGCTGACCCCGCGCGAAC 510  
 Db 109 AspTyriIleHisGluGlnAspIleAsnCysPheAsnSerAlaLeu----- 123  
 QY 511 CTGTCAAGAAGAGCTGGAAGCGCCCAACAGAGCGCACTTTTCCCTGCGAATGAAGAGC 570  
 Db 124 -----AsnTyCysAspLeuAsnTrpProGlnMetCysAsnValArgValLysSer 140  
 QY 571 ACCTGACACAGAGCGCGCGCTCAACTCAAGCGGCCACC---TGAAGGTGCTG 627  
 Db 141 SerLeuThrLysArgAlaAsnLysAspAlaValArgAlaSerProGlyTyriLysValLeu 160  
 QY 628 CACTGCTCAGGACATATGAGCGCTCAACAGC-----CCCTGTCACAGACT 672  
 Db 161 ArgLeuGluIleThrMet-GlyProAsnThrAsnThrArgMetIleAlaCysTyPrMe 180  
 QY 673 TCCCTGCGGAGCGCTCGCTCGGAGCCTCCCTGCAATGCCCTGGTGCTTA 724  
 Db 180 tProThrProValLeuSerThrValThrIleProSerAsnSerPheValIle 197

RESULT 12  
 JC7633  
 aryl hydrocarbon nuclear translocator ARNT2-like factor, ARNT2X - zebra fish  
 C;Species: Brachydanio rerio (zebra fish)  
 C;Date: 30-Jun-2001 #sequence\_revision 30-Jun-2001 #text\_change 30-Jun-2001  
 C;Accession: JC7633  
 R;Hsu, H.J.; Wang, W.D.; Hu, C.H.  
 Biochem. Biophys. Res. Commun. 282, 487-492, 2001  
 A;Title: Ectopic expression of negative ARNT2 factor disrupts fish development.  
 A;Reference number: JC7633; PMID:21294759; PMID:11401485  
 A;Contents: Heart  
 A;Accession: JC7633  
 A;Molecule type: mRNA  
 A;Residues: 1-392 <HSU>  
 A;Cross-references: GB:AY007992  
 C;Comment: This factor, a basic helix-loop-helix PAS (bHLH-PAS) factor, heterodimerizes  
 C;Genetics:  
 A;Gene: arnt2X

Alignment Scores: 2.5e-05 Length: 392  
 Pred. No.:

Score: 166.50 Matches: 62  
 Percent Similarity: 40.43% Conservative: 50  
 Best Local Similarity: 22.38% Mismatches: 83  
 Query Match: 8.28% Indels: 83  
 DB: 2 Gaps: 9

US-09-896-791B-2 (1-1100) x JC7633 (1-392)

QY 64 AAGGAGAAGTCGCGGACGCGCGCGCGCGAGCGAGGAGTGCCTGTAC 123  
 Db 47 ArgGluAsnHisSerGluIleGluArgArgArgAsnLysMetThrGlnTyriIleThr 66  
 QY 124 CAGCTGGCGCACACTCTGCCCTTTGGCGCGGTCTGAGCGCGACCTGGACAAGCGCTCC 183  
 Db 67 GluLeuSerAspMetValProThrCysSerAlaLeuAlaArgLysProAspLysLeuThr 86  
 QY 184 ATCATGCGCTCACATCAGTACCTACCTGCGCATGACCGCTCTGCGCAGCAGGTGAAAA 243  
 Db 87 IleLeuArgMetAlaValSerHisMetLysSerMetArg-----Gly-Th 101  
 QY 244 AGGGGAGAGCCCACTGAGCGCTCTGCTAC----- 271  
 Db 101 rGlyAsnThrSerThrAspGlyAlaTyLysProSerPheLeuThrGluGlnGluLeu 121  
 QY 272 -----CTGAAGCGCCTGGAGGTTTCGTACTGTTACTCACCCTCGAG---GGAGA 318  
 Db 121 sHisLeuIleLeuGluAlaAlaAspGlyPheLeuPheValAlaAlaGluThrGlyAr 141  
 QY 319 CATGGCTTACCTCTCGGAAATCTCAGCAAGCACCTGGG---CCTCAGTCAGTGGACCT 374  
 Db 141 gValIleTyriValSerAspSerValThrProValLeuAsnHisProGlnSer----- 158  
 QY 375 CTGTTCTCTCTCCCTGATACATAACCCCACTCTGTTGTTACCAATTTCTCTCTGGAGCTCAT 434  
 Db 159 -----GluTrpPh 161  
 QY 435 TGGACACATATCTTTGATTTTATCCATCCTGTGACCAAGAGAACTTCAAGACGCGCT 494  
 Db 161 eGlySerThrLeuPheGluGlnValHisProAspValAspLysLeuArgGluGlnLe 181  
 QY 495 GACCCCGAGCGCCCACTGTCAAGAAGAAGCTG----- 528  
 Db 181 uSerThrSerGluAsnSerMetThrGlyArgIleLeuAspLeuLysThrGlyThrVal 201  
 QY 528 ----- 528  
 Db 201 sLysGluGlyGlnSerSerMetArgMetCysMetGlySerArgArgSerPheIleCy 221  
 QY 529 -----GAAGCCCAACAGAGCGCCCACTTTTCCCTGCGAATGAAGAGCAC 572  
 Db 221 sArgMetArgCysGlySerAlaProLeuAsp---HisIleSerLeuAsnArgLeuSer 240  
 QY 573 GCTCACCAGCAGAGGCGGCACG-----CTCAACCTCAAGCGCCCACTGGAA 620  
 Db 240 rMetArgLysArgTyArgAsnGlyLeuGlyProSerLysGluGlyGluAlaGlnTy 260  
 QY 621 GGTGCTGCATGCTCAGGACATATGAGGCGCTACAGCCCTCCAGCCCTTCCCTCCCT-- 678  
 Db 260 rValValHisCysThrGlyTyriLysAlaTrpProAlaGlyMetThrIleProAs 280  
 QY 679 -GCCGGAGCGCTCGCTCCGAGCGCTCCCTGCAATGCTGCTGCTTATC 726  
 Db 280 pGluAspThrGluAlaGlyGlnThrSerLysTyriCysLeuValAlaIle 296

RESULT 13  
 JC7635  
 aryl hydrocarbon receptor nuclear translocator 1 - chicken  
 N;Alternate names: hypoxia-inducible factor 1 beta  
 C;Species: Gallus gallus (chicken)  
 C;Date: 30-Jun-2001 #sequence\_revision 30-Jun-2001 #text\_change 30-Jun-2001  
 C;Accession: JC7635  
 R;Catron, T.; Mendiola, M.A.; Smith, S.M.; Born, J.; Walker, M.K.  
 Biochem. Biophys. Res. Commun. 282, 602-607, 2001

A:Title: Hypoxia regulates avian cardiac Arnt and HIF-1alpha mRNA expression.  
A:Reference number: JC7635; MUID:21294777; PMID:11401503

A:Contents: Embryo

A:Accession: JC7635

A:Molecule type: mRNA

A:Residues: 1-805 <CAT>

C:Comment: This protein, known as hypoxia-inducible factor 1 beta, a member of the basic helix-loop-helix family, is a transcription factor that binds to hypoxia-inducible DNA sequences to mediate responses to hormonal and environmental stimuli. It is also a common dimer partner for transcription factors such as p53, Jun, Fos, and Myc.

C:Genetics:

A:Gene: arntl

C:Keywords: transcription factor

Alignment Scores:  
Pred. No.: 5,81e-05 Length: 805  
Score: 162.00 Matches: 66  
Percent Similarity: 36.25% Conservative: 46  
Best Local Similarity: 21.36% Mismatches: 76  
Query Match: 8.06% Indels: 122  
DB: 2 Gaps: 9

US-09-896-791B-2 (1-1100) x JC7635 (1-805)

```
QY 58 CTGCGGAGGAGAGTCGGGGACGGCGCCGACGGCGGCGGAGGAGGAGG 117
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 89 LeuAlaArgGluAsnHisSerGluIleGluArgArgArgAsnLysMetThrAlaTyr 108
QY 118 CTGTACCACTGGCGACACTTCGCCCTTTCGGCGCGGCTGAGCGGCGACCTGACACAG 177
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 109 IleThrGluLeuSerAspMetValProThrCysSerAlaLeuAlaArgLysProAspLys 128
QY 178 GCCTCCATCATCGGCTCAACATCAGTACCTACCTGCGGATCGCGCTCTGCGGACGAGT 237
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 129 LeuThrIleLeuArgMetAlaValSerHisMetLysSerLeuArg 143
QY 238 GGAAGAAGGGGAGAGCCACTGAGCCCTGCTAC----- 271
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 144 Gly-ThrGlyAsnThrSerThrAspGlyThrTyrLysProSerPheLeuThrAspGlnG1 163
QY 272 -----CTGAAGCCCTGAGGAGTTCGTCATGCTACTCACCAGCGAG-- 313
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 163 uLeuLysHisLeuIleLeuGluAlaAlaAspGlyPheLeuPheIleValSerCysGluTh 183
QY 314 -GGACATGGCTTACCTCGGAAATGTACGACAGACCTGGG-----CCTCAGTCACT 368
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 183 rGlyArgValValTyrValSerAspSerValThrProValLeuAsnGlnProGlnSer-- 202
QY 369 GGACCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTGGA 428
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 203 -----G1 203
QY 429 GCTCATTGGACACAGTATCTTTGATTTTATCCATCCCTGATGGTACTCACCAGGAGTCAAGA 488
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 203 utrPheGlySerThrLeuTyrAspGlnValHisProAspAspValGlyLysLeuArgG1 223
QY 489 CGCCTGACCCCGAGCGCAAC-----CTGTCAAA 518
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 223 uGlnLeuSerThrSerGluAsnAlaLeuThrGluGlyThrLysProThrPcysLeuSerAs 243
QY 519 GAAGAGCTGGAGCCCAACAGAGCGCCACTTTTCCCTGCGAATGAAGACGCGCTCAC 578
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 243 nLysAspProAlaAlaProGluSer-----A1 253
QY 579 CAGCAGAGGGCGGACGCTCAACTCAAGCGGCGGACCTGGAG----- 621
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 253 aSerLysGlyArgIleLeuAspLeuLysThrGlyThrValLysLysGluGlyGlnGlnSe 273
QY 621 ----- 621
Db 273 rMetArgMetCysMetGlySerArgArgSerPheIleCysArgMetArgCysGlyAsnSe 293
QY 621 ----- 621
Db 293 rSerValAspAlaValSerValAsnArgLeuSerPheMetArgAsnArgCysArgAsnG1 313
```

```
QY 622 -----GTGCTGCACCTGCTCAGACATAT 644
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 313 yLeuGlyAlaThrLysAspGlyLupProHisTyrValValValHisCysThrGlyTyrI1 333
QY 645 GAGGCGCTACAAAGCCCTGCACAGACTTCCCTCCCGGAGCCCT---CGCTCCGAGCC 701
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 333 elysAlaTrpProAlaGlyValSerLeuProAspAspProAspAlaGlyGlnG1 353
QY 702 TCCCTGCAATGCCCTGGTCTTATC 726
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 353 ySerLysPheCysLeuValAlaIle 361
RESULT 14
A56241
aryl hydrocarbon receptor nuclear translocator protein - mouse
C:Species: Mus musculus (house mouse)
C:Date: 03-Oct-1995 #sequence_revision 03-Oct-1995 #text_change 05-Nov-1999
C:Accession: A56241
R:Reisz-Porszasz, S.; Probst, M.R.; Fukunaga, B.N.; Hankinson, O.
Mol. Cell. Biol. 14, 6075-6086, 1994
A:Title: Identification of functional domains of the aryl hydrocarbon receptor nuclea
A:Reference number: A56241; MUID:94344118; PMID:8065341
A:Accession: A56241
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-791 <RET>
A:Cross-references: GB:U10325; NID:9555687; PIDN:AA56717.1; PID:9555688
A:Note: authors translated the codon TTT for residue 375 as Ser, TGT for residue 380
A:Note: authors failed to translate CAG for residue 507 as Gln
C:Genetics:
A:Gene: ARNT
```

Alignment Scores:  
Pred. No.: 0.000111 Length: 791  
Score: 158.50 Matches: 63  
Percent Similarity: 41.73% Conservative: 53  
Best Local Similarity: 22.66% Mismatches: 82  
Query Match: 7.89% Indels: 81  
DB: 2 Gaps: 10

US-09-896-791B-2 (1-1100) x A56241 (1-791)

```
QY 58 CTGCGGAGGAGAGTCGGGGACGGCGCCGCGGCGGAGGAGGAGG 117
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 89 LeuAlaArgGluAsnHisSerGluIleGluArgArgArgAsnLysMetThrAlaTyr 108
QY 118 CTGTACCACTGGCGACACTCTGCCCTTTCGGCGCGGCTGAGCGGCGACCTGGAACAAG 177
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 109 IleThrGluLeuSerAspMetValProThrCysSerAlaLeuAlaArgLysProAspLys 128
QY 178 GCCTCCATCATCGGCTCAACATCAGTACCTACCTGCGGATCGCGCTCTGCGGACGAGT 237
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 129 LeuThrIleLeuArgMetAlaValSerHisMetLysSerLeuArg 143
QY 238 GGAAGAAGGGGAGAGCCACTGAGCCCTGCTAC----- 271
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 144 Gly-ThrGlyAsnThrSerThrAspGlyThrTyrLysProSerPheLeuThrAspGlnG1 163
QY 272 -----CTGAAGCCCTGAGGAGTTCGTCATGCTACTCACCAGCGAG-- 313
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 163 uLeuLysHisLeuIleLeuGluAlaAlaAspGlyPheLeuPheIleValSerCysGluTh 183
QY 314 -GGACATGGCTTACCTCGGAAATGTACGACAGACCTGGG-----CCTCAGTCACT 368
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 183 rGlyArgValValTyrValSerAspSerValThrProValLeuAsnGlnProGlnSer-- 202
QY 369 GGACCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTGGA 428
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 203 -----G1 203
QY 429 GCTCATTGGACACAGTATCTTTGATTTTATCCATCCCTGATGGTACTCACCAGGAGTCAAGA 488
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

D	b	203	utrPheGlySerThrLeuTyrAspGlnValHisProaspaspValAspLysLeuArgGl	223
Q	y	489	CGCCCTGACCCCCAGGCCGAAC---CTGTCAAGAAGAAAGCTGGGAAGCCCCAACAA-	540
	:		:    :::        ::: :::    :::	
D	b	223	uGlnLeuSerThrSerGluasnAlalaLeuThrGlyArgValLeuAspLeuLysThrGlyTh	243
Q	y	541	-----GAGCGGCACATT	551
D	b	243	rValLysLysGluGlycInlSerSerMetArgMetCysMetGlySerArgArgSerPh	263
Q	y	* 552	TTCCCTCGCAATGAAGACGCGTCACCAGC-::   ::	582
D	b	263	eIleCysArgMetArgCysGlyThrSerSerValaspProValSerMetasnArgLeuSe	283
Q	y	* 583	-----AGAGGGCGCACGCTCAACTCAAACGGCCACC-----TG	617
	:		:             :::	::
D	b	283	rPheLeuArgAsnArgCysArgasnglyLeuGlySerValLysGluGlyGluProHisPh	303
Q	y	618	GAGGTGCTCAGCTGCTCAGGCATATGAGGCGCTACAGGCCCTGCACAGACTTCCC	677
	:		:    :::   :::   :::   :::   :::	::
D	b	303	eValValHisCysThrGlyTyrIleLysalatrProAlaGlyValserLeuPr	323
Q	y	678	TGCCGGGAGCGCTCGCTCC---GAGCGTCCCTCCGTAATGCTGTGGCTATC	726
	:		:     :::: ::::	::
D	b	323	oAspAspProglualaglycInglySerLysPhecysLeuAlalle	340
			RESULT 15	
			JCS5405	
			brain and muscle Ah receptor nuclear translocator-like protein, BMAL1b - human	
			C:Species: Homo sapiens (nan)	
			C>Date: 10-Jun-1997 #sequence_revision 18-Jul-1997 #text_change 01-Dec-2000	
			C:Accession: JCS5405; JC5404	
			R:Ikeda, M.; Nomura, M.	
			Biochem. Biophys. Res. Commun. 233, 258-264, 1997	
			A:Title: cDNA cloning and tissue-specific expression of a novel basic helix-loop	
			site usage.	
			A:Reference number: JCS5404; MUID:97289529; PMID:9144434	
			A:Accession: JCS5405	
			A>Status: nucleic acid sequence not shown	
			A:Molecule type: mRNA	
			A:Residues: 1-626 <IK2>	
			A:Cross-references: DDBJ:AB000812; NID:g2094734; PIDN:BAA19935.1; PID:g2094735	
			A:Accession: JCS5404	
			A>Status: nucleic acid sequence not shown	
			A:Molecule type: mRNA	
			A:Residues: 'MINI',48-415,'P',417-626 <IK2>	
			A:Cross-references: DDBJ:D89722	
			C:Comment: This protein plays a role in the control of central nervous system	
			C:Genetics:	
			A:Introns: 47/2; 223/1; 250/2; 274/1; 288/2; 361/1; 392/1; 442/2	
			Alignment Scores:	
			Pred. No.: 0.000159 Length: 626	
			Score: 156.50 Matches: 87	
			Percent Similarity: 39.94% Conservative: 54	
			Best Local Similarity: 24.65% Mismatches: 110	
			Query Match: 7.79% Indels: 102	
			DB: 2 Gaps: 16	
			US-09-896-791B-2 (1-1100) x JCS5405 (1-626)	
Q	y	48	GACACCGAGCTCGGAAGAGAGAGTCCGGGACCGG-----GCCCGACCGC	95
	:		:      :::     :::: ::   :::	::
D	b	65	GLuHIScIngly-ArGileLysasnAlaArGluahHisSerGlnlleGluLysArGAr	84
Q	y	96	CGCGACGACGAGGAGCGAGGTGCTGTACCGAGCTGGCGCACACTCTGCCTCTTGGCGCGG	155
	:		:      :::: :::: ::   :::	::
D	b	84	gArgAspLysMetAsnSerPheilleaspGluLeualaSerLeuValProThrCysasnAl	104
Q	y	156	CGTCAGCGCGCACCTCGACAAGCGCTCCATCATCGCGCTCAACATCAGCTACCTACCTCGCAT	215
	:		: :                ::   :::   :::   :::	::
D	b	104	aMetSerArgLysLeuLysLeuThrValLeuArqMetaValGlnHisMetArqTh	124

Search completed: December 2, 2002, 20:40:48  
Job time : 49 secs

—



GenCore version 5.1.3  
Copyright (c) 1993 - 2002 Compugen Ltd.

OM nucleic - protein search, using frame\_plus\_n2p model

Run on: December 2, 2002, 19:48:55 ; Search time 23 Seconds

(without alignments)  
3967.301 Million cell updates/sec

Title: US-09-896-791B-2

Perfect score: 2010

Sequence: 1 gaattcgccagcaggccat.....aaaaaaaaaacatgcggccgc 1100

Scoring table: BLOSUM62

Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 225784

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+n2p.model -DEV=xlp  
-Q/cgn2\_1/USPTO\_spool/US09896791/runat\_02122002\_152658\_5574/app\_query.fasta\_1.1287  
-DB=SwissProt\_40 -QFMT=fastan -SUFFIX=rs -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0  
-UNITS=bits -START=1 -END=1 -MATRIX=biosum62 -TRANS=human40.cdi -LIST=45  
-DOALIGN=200 -THR\_SCORE=pct -THR\_MAX=100 -THR\_MIN=0 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US09896791.ecgn\_1\_1\_13.@runat\_02122002\_152658\_5574 -NCPU=6 -ICPU=3  
-NO\_XLPXY -NO\_MMAP -LARGESQUERY -NEG\_SCORES=0 -WAIT -LONGLOG -DEV\_TIMEOUT=120  
-WARN\_TIMEOUT=30 -THREDS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOPOP=6 -FGAPEXT=7  
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : SwissProt\_40.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	613.5	30.5	822	1 HIFA_MOUSE	Q61221 mus musculus
2	606.5	30.2	826	1 HIFA_HUMAN	Q16665 homo sapien
3	593.5	29.5	874	1 PAS1_MOUSE	P97481 mus musculus
4	572.5	28.5	870	1 PAS1_HUMAN	Q99814 homo sapien
5	346	17.2	1507	1 SIMA_DROME	Q24167 drosophila
6	340.5	16.9	590	1 NPAL_HUMAN	Q99742 homo sapien
7	337	16.8	594	1 NPAL_MOUSE	P97459 mus musculus
8	324	16.1	657	1 SIM2_MOUSE	Q61079 mus musculus
9	312.5	15.5	766	1 SIM1_HUMAN	P81133 homo sapien
10	311	15.5	667	1 SIM2_HUMAN	Q14190 homo sapien
11	310.5	15.4	765	1 SIM1_MOUSE	Q61045 mus musculus
12	302.5	15.0	958	1 TRH_DROME	Q24119 drosophila
13	289.5	14.4	673	1 SIM_DROME	P05709 drosophila
14	165.5	8.2	712	1 ARN2_MOUSE	Q61324 mus musculus
15	165	8.2	706	1 ARN2_HUMAN	Q9Hb22 homo sapien
16	158.5	7.9	413	1 CYCL_DROME	Q61734 drosophila
17	158.5	7.9	791	1 ARNT_MOUSE	P53762 mus musculus
18	157.5	7.8	800	1 ARNT_HUMAN	P41739 rattus norv

19	156.5	7.8	789	1 ARNT_HUMAN	P27540 homo sapien
20	155.5	7.7	790	1 ARNT_RABIT	O02748 oryctolagus
21	153.5	7.6	583	1 BMAL_HUMAN	O00327 homo sapien
22	145	7.2	848	1 AHR_HUMAN	P35869 homo sapien
23	140	7.0	805	1 AHR_MOUSE	P30561 mus musculus
24	138.5	6.9	853	1 AHR_RAT	P41738 rattus norv
25	135.5	6.7	1097	1 CCT_DROME	O96433 drosophila
26	130.5	6.5	855	1 CLOC_MOUSE	O08785 mus musculus
27	129.5	6.4	644	1 ARNT_DROME	O15945 drosophila
28	128.5	6.4	1464	1 NCO2_HUMAN	O15596 homo sapien
29	126.5	6.3	846	1 CLOC_HUMAN	O15516 homo sapien
30	126	6.3	816	1 NPA2_MOUSE	P97460 mus musculus
31	126	6.3	937	1 CDA7_HUMAN	Q9un72 homo sapien
32	126	6.3	1462	1 NCO2_MOUSE	Q61026 mus musculus
33	125.5	6.2	1465	1 NCO2_RAT	Q9wu19 rattus norv
34	121.5	6.0	717	1 PRDD_HUMAN	Q9H4q3 homo sapien
35	121.5	6.0	1183	1 DRPL_RAT	P54258 rattus norv
36	121	6.0	824	1 NPA2_HUMAN	Q99743 homo sapien
37	120	6.0	1516	1 NCO2_XENLA	Q9W705 xenopus lae
38	119	5.9	950	1 CDA1_HUMAN	Q9V513 homo sapien
39	119	5.9	3530	1 MX15_HUMAN	Q9ukn7 homo sapien
40	118.5	5.9	215	1 Y215_ADE02	P03291 human adeno
41	117.5	5.8	649	1 V70K_EPMV	P20129 eggplant mo
42	117	5.8	948	1 CDA2_HUMAN	Q9Y5h9 homo sapien
43	116.5	5.8	640	1 ELI2_HUMAN	O00472 homo sapien
44	116.5	5.8	941	1 CDAC_HUMAN	Q9un75 homo sapien
45	114.5	5.7	443	1 GAT3_HUMAN	P23771 homo sapien

#### ALIGNMENTS

RESULT 1

HIFA\_MOUSE

ID HIFA\_MOUSE STANDARD; PRT: 822 AA.  
AC Q61221; Q61665; Q61664; O08993; O08741;  
DT 01-NOV-1997 (Rel. 35, Created)  
DT 15-DEC-1998 (Rel. 37, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Hypoxia-inducible factor 1 alpha (HIF-1 alpha) (ARNT interacting protein).  
GN HIF1A.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sclurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=C57BL/6; TISSUE=Hepatocytes;  
RX MEDLINE=96355491; PubMed=8702901;  
RA Li H., Ko H.P., Whitlock J.P. Jr.;  
RT "Induction of phosphoglycerate kinase 1 gene expression by hypoxia.  
RT Roles of Arnt and Hif1alpha".  
RL J. Biol. Chem. 271:21262-21267(1996).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=129/SvJ;  
RX MEDLINE=98034461; PubMed=9368100;  
RA Luo G., Gu Y.-Z., Jain S., Chan K.M., Hogenesch J.B.,  
RA Bradfield C.A.;  
RT "Molecular characterization of the murine Hif-1 alpha locus".  
RL Gene Expr. 6:287-299(1997).  
RN [3]  
RP SEQUENCE FROM N.A.  
RC STRAIN=129/SvJ;  
RX MEDLINE=97354184; PubMed=9210478;  
RA Wenger R.H., Rolfs A., Kvietikova I., Spielmann P., Zimmermann D.R.,  
RA Gassmann M.;  
RT "The mouse gene for hypoxia-inducible factor-1alpha. Genomic organization, expression and characterization of an alternative first exon and 5' flanking sequence".  
RL Eur. J. Biochem. 246:155-165(1997).  
RN [4]  
RP SEQUENCE OF 13-822 FROM N.A.



QY 739 CCC 741  
 Db 228 Pro 228  
 RESULT 2  
 HIFA\_HUMAN STANDARD; PRT; 826 AA.  
 AC Q16665;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Hypoxia-inducible factor 1 alpha (HIF-1 alpha) (ARNT interacting  
 protein) (Member of PAS protein 1) (MOP1) (HIF1 alpha).  
 OS HIFA.  
 GN Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 CC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A., AND SEQUENCE OF 166-170; 259-289 AND 771-781.  
 RX MEDLINE=95296340; PubMed=7539918;  
 RA Wang G.L., Jiang B.H., Rue E.A., Semenza G.L.;  
 RT "Hypoxia-inducible factor 1 is a basic-helix-loop-helix-PAS  
 heterodimer regulated by cellular O2 tension.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 92:5510-5514(1995).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Hepatoma;  
 RX HOGENSECH J.B., Chan W.K., Jackiw V.H., Brown R.C., Gu Y.-Z.,  
 RA Pray-Grant M., Perdew G.H., Bradfield C.A.;  
 RT "Characterization of a subset of the basic-helix-loop-helix-PAS  
 superfamily that interacts with components of the dioxin signaling  
 pathway".  
 RL J. Biol. Chem. 272:8581-8593(1997).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RA Rupert J.L., Hochachka P.W.;  
 RT "Hif1a sequence in the Quechua, a high altitude population.";  
 RL Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.  
 CC -!- FUNCTION: INVOLVED IN THE INDUCTION OF OXYGEN REGULATED GENES.  
 CC SPECIFICALLY RECOGNIZES AN 8 BP HYPOXIA RESPONSE ELEMENT (HRE).  
 CC -!- SUBUNIT: EFFICIENT DNA BINDING REQUIRES DIMERIZATION WITH ANOTHER  
 CC BHLH PROTEIN. HETERODIMER OF AN ALPHA AND A BETA (ARNT) SUBUNITS.  
 CC THE ALPHA SUBUNIT IS UNIQUE TO HIF-1 WHEREAS HIF-1 BETA (ARNT) CAN  
 CC DIMERIZE WITH OTHER BHLH-PAS PROTEINS. INTERACTS WITH HSP90.  
 CC -!- SUBCELLULAR LOCATION: Nuclear (Potential).  
 CC -!- TISSUE SPECIFICITY: EXPRESSED IN MOST TISSUES. HIGHEST LEVELS IN  
 CC KIDNEY AND HEART.  
 CC -!- INDUCTION: UNDER REDUCED OXYGEN TENSION.  
 CC -!- DOMAIN: BOTH HYPOXIA RESPONSIVENESS AND TRANSACTIVATION CAPABILITY  
 CC RESIDE WITHIN THE C-TERMINAL PART.  
 CC -!- PTM: REQUIRES PHOSPHORYLATION FOR DNA-BINDING.  
 CC -!- SIMILARITY: BELONGS TO THE BASIC HELIX-LOOP-HELIX (BHLH) FAMILY OF  
 CC TRANSCRIPTION FACTORS.  
 CC -!- SIMILARITY: CONTAINS 2 PAS (PER-ARNT-SIM) DIMERIZATION DOMAINS.  
 CC -!- SIMILARITY: CONTAINS 1 PAS-ASSOCIATED C-TERMINAL (PAC) DOMAIN.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL; U22431; AAC50152.1; -  
 CC EMBL; U29165; AAC51210.1; -  
 CC EMBL; AF207601; AAF20139.1; -  
 CC EMBL; AF207602; AAF20140.1; -  
 CC EMBL; AF208487; AAF20149.1; -  
 CC TRANSFAC; T01610; -  
 CC Genbank; HGNC:4910; HIF1A.

DR MIM: 603348; -  
 DR InterPro: IPR001092; HLH\_basic.  
 DR InterPro: IPR001321; Hypoxindf1A.  
 DR InterPro: IPR001610; PAC.  
 DR InterPro: IPR000014; PAS\_domain.  
 DR Pfam: PF00785; PAC; 1.  
 DR Pfam: PF00989; PAS; 2.  
 DR PRINTS: PR01080; HYPOXIAIFIA.  
 DR SMART: SM00353; HLH; 1.  
 DR SMART: SM00086; PAC; 1.  
 DR SMART: SM00091; PAS; 2.  
 DR PROSITE: PS00038; HLH\_1; 1.  
 DR PROSITE: PS00888; HLH\_2; 1.  
 DR PROSITE: PS0112; PAS; 2.  
 DR Repeat: DNA-binding; Nuclear protein; Transcription regulation;  
 KW Activator; Phosphorylation.  
 FT DNA\_BIND 17 30 BASIC DOMAIN.  
 FT DOMAIN 31 71 HELIX-LOOP-HELIX MOTIF (BY SIMILARITY).  
 FT DOMAIN 85 158 PAS 1.  
 FT DOMAIN 228 298 PAS 2.  
 FT DOMAIN 302 345 PAC.  
 FT DOMAIN 615 621 POLY-THR.  
 SQ SEQUENCE 826 AA; 92670 MW; ABD4F7DAA135BE2D CRC64;  
 Alignment Scores:  
 Pred. No.: 2.57e-35 Length: 826  
 Score: 606.50 Matches: 129  
 Percent Similarity: 68.07% Conservative: 33  
 Best Local Similarity: 54.20% Mismatches: 48  
 Query Match: 30.17% Indels: 28  
 Db: 1 Gaps: 3  
 US-09-896-791B-2 (1-1100) x HIFA\_HUMAN (1-826)  
 QY 49 AACACCGAGTGGCGAAGAGAAAGTGGCGGGACGCGCGCGCGAGCGGCGAGCAGGAG 108  
 Db 14 SerSerGluArgArgLysGluSerArgAspAlaAlaArgSerArgArgSerLysGlu 33  
 QY 109 ACGGAGGTGCTGCTACCGAGTGGCGGACACTCTGCGCCTTTGCGCGGGGCTGACGCGGAC 168  
 Db 34 SerGluValPheTyrGluLeuAlaHisGlnLeuProLeuProHisAsnValSerHis 53  
 QY 169 CTGGCAAGGCTCCATCGCTCAATCAGTCACTGCTGCGGAGTGGCGGAGTGGCGGAGTGGC 228  
 Db 54 LeuAspLysAlaSerValMetArgLeuThrIleSerTyrLeuArgValArgLysLeuLeu 73  
 QY 229 GCAGCAGTGGGAAAGGGGGAGAG-----CCACTGGAGCGCTGCTACT 273  
 Db 74 AspAlaGly-AspLeuAspIleGluAspMetLysAlaGlnMetAsnCysPheTyrIle 93  
 QY 274 GAAGCGCTGGAGGTTTCGTCATGTCACCGCGGAGGAGGAGCATGGCTTACTGTC 333  
 Db 93 uLysAlaLeuaspGlyPheValMetValLeuThrAspAspGlyAspMetIleTyrIle 113  
 QY 334 GAAATATGTCAAGACGACCTGGCGCTCAGTCAGTGGAGGAGCTGTCTCTCTCTCTCTCTCTCT 393  
 Db 113 rAspAsnValAsnLysTyrMetGlyLeuThrGln----- 124  
 QY 394 CATAACCCCATCTCTGGTACCACATTTCTCTCTGGAGCTCATTTGGACACAGTATCTTTGAT 453  
 Db 125 -----Phe--GluLeuThrGlyHisSerValPheAsp 134  
 QY 454 TTTATCCATCCCTGTGACCAAGAGCAACTTCAAGAGCGCCTCAGCCCGGAGGCGGACCTG 513  
 Db 135 PheThrHisProCysAspHisGluMetArgGluMetLeuThrHisArgAsnGlyLeu 154  
 QY 514 TCAAGAAGAAGCTGGAAGCCCAACAGACGCGCCACTTTTCTCTCGAATGAAGACGACG 573  
 Db 155 ValLysLysGlyLysGluGlnAsnThrGlnArgSerPhePheLeuArgMetLysCysThr 174  
 QY 574 CTCACACGAGGCGGCGACCGCTCAACCTCAAGGGCGGACCTGGAAGGTGCTGCACTGCG 633  
 Db 175 LeuThrSerArgGlyArgThrMetAsnIleLysSerAlaThrTrpLysValLeuHisCys 194



```

Db 41 LeuProLeuProHisSerValSerHisLeuAspLysAlaSerIleMetArgLeuAla 60
|||||
199 ATCAGCTACCTGCGCATCGCCGCTC-----TGGCAGCAGGTGGAAAAGG 246
|||||
61 IleSerPheLeuArgThrHisLeuLeuSerValCysSerGluAsnGluSerG1 80
|||||
247 GGGAG-----CCACTGGAGCGCTGCTACCTGAAGCCCTCGAGGGTTTCGTGAT 297
|||||
80 uAlaGluAlaAspGlnGlnMetAspAsnLeuTyrLeuLysAlaLeuGluGlyPheIleAl 100
|||||
298 GGTACTACCGCGGAGGAGACATGCTTACTGTGGAAATGTCACAGACCTGGG 357
|||||
100 aValValThrGlnAspGlyAspMetIlePheLeuSerGluAsnIleSerLysPheMetG1 120
|||||
358 CTTAGTCAAGTGGAGCTCTGTTCTCCCTCCCTGATACATAACCCCACTCCTGGTACCAAT 417
|||||
120 yLeuThrGln----- 123
Qy 418 TTCCTCTGGAGCTCATTTGGACACAGTATCTTTGATTTTCCATCCCTGTGACCAAGAG 477
|||
124 ---Val--GluLeuThrGlyHisSerIlePheAspPheThrHisProCysAspHisGlu 141
|||||
478 GAACCTCAAGAGCGCTCACCCCGAG-----CGAACCTGTCTCAAGAAGAAGCTGGAA 531
|||||
142 GluIleArgGluAsnLeuThrLeuLysAsnGlySerGlyPheGlyLysSerLysAsp 161
|||||
532 GCCCAACAGAGCGGCCTTTCTCCCTGCGAATGAAGACGCGCTCACCGAGCGGCGC 591
|||||
162 ValSerThrGluArgAspPhePheMetArgMetLysCysThrValThrAsnArgGlyArg 181
|||||
592 ACGCTCAACCTCAAGCGCGCCACCTGGAAGTGTGCTGCTGCTCAGGACATATGAGGCC 651
|||||
182 ThrValAsnLeuLysSerAlaThrTrpLysValLeuHisCysThrGlyGlnValArgVal 201
|||||
652 TACAAAGCCC---CCTGCACAGACTTCCCTCGCGGAGCCCTCGCTCGAGCCTCCCTCG 708
|||||
202 TyrAsnAsnCysProHisSerSerLeuCysGlySer-----LysGluProLeuLeu 219
|||||
709 CAATGCTGCTGCTATCTGTGAAGCCATCCCGCCAGCTCCCTCCCTCCAC 756
|||||
220 SerCysLeuIleIleMetCysGluProIle---GlnHisProSerHis 234
|||||
RESULT 4
PAS1_HUMAN STANDARD; PRT; 870 AA.
AC Q9814; Q98630;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DE 15-JUN-2002 (Rel. 41, Last annotation update)
DE Endothelial PAS domain protein 1 (EPAS-1) (Member of PAS protein 2)
DE (MOP2).
DE PAS1.
DE Homo sapiens (Human).
OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97152468; PubMed=9000051;
RA Tian H., McKnight S.L., Russell D.W.;
RT "Endothelial PAS domain protein 1 (EPAS1), a transcription factor
selectively expressed in endothelial cells.";
RL Genes Dev. 11:72-82(1997).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Hepatoma;
RX MEDLINE=97236817; PubMed=9079689;
RA Hogenesch J.B., Chan W.K., Jackiw V.H., Brown R.C., Gu Y.-Z.,
RA Pray-Grant M., Perdew G.H., Bradford C.A.;
RT "Characterization of a subset of the basic-helix-loop-helix-PAS
superfamily that interacts with components of the dioxin signaling
pathway.";
```

```

J. Biol. Chem. 272:8581-8593(1997).
-|- FUNCTION: TRANSCRIPTION FACTOR INVOLVED IN THE INDUCTION OF OXYGEN
REGULATED GENES. SPECIFICALLY RECOGNIZES AN 8 BP HYPOXIA RESPONSE
ELEMENT (HRE). REGULATES THE VASCULAR ENDOTHELIAL GROWTH FACTOR
(VEGF) EXPRESSION AND SEEMS TO BE IMPLICATED IN THE DEVELOPMENT OF
BLOOD VESSELS AND THE TUBULAR SYSTEM OF LUNG. MAY ALSO PLAY A ROLE
IN THE FORMATION OF THE ENDOTHELIUM THAT GIVES RISE TO THE BLOOD
BRAIN BARRIER. POTENT ACTIVATOR OF THE TIE-2 TYROSINE KINASE
EXPRESSION.
-|- SUBUNIT: EFFICIENT DNA BINDING REQUIRES DIMERIZATION WITH ANOTHER
BHLH PROTEIN. HETERODIMER WITH THE ARNT PROTEIN.
-|- SUBCELLULAR LOCATION: Nuclear (Potential).
-|- TISSUE SPECIFICITY: EXPRESSED IN MOST TISSUES, WITH HIGHEST LEVELS
IN PLACENTA, LUNG AND HEART. SELECTIVELY EXPRESSED IN ENDOTHELIAL
CELLS.
-|- SIMILARITY: BELONGS TO THE BASIC HELIX-LOOP-HELIX (BHLH) FAMILY OF
TRANSCRIPTION FACTORS. HIGHEST TO HIF-1 ALPHA.
-|- SIMILARITY: CONTAINS 2 PAS (PER-ARNT-SIM) DIMERIZATION DOMAINS.
-|- SIMILARITY: CONTAINS 1 PAS-ASSOCIATED C-TERMINAL (PAC) DOMAIN.
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----
EMBL; U81984; AAB41495.1; -;
DR EMBL; U51626; AAC51212.1; -;
DR TRANSFAC; T02718; -;
DR Genew; HGNC:3374; EPAS1.
DR MIN; 603349; -;
DR InterPro; IPR001092; HLH_basic.
DR InterPro; IPR001610; PAC.
DR InterPro; IPR000014; PAS_domain.
DR Pfam; PF00785; PAC; 1.
DR Pfam; PF00989; PAS; 2.
DR SMART; SM00353; HLH; 1.
DR SMART; SM00086; PAC; 1.
DR SMART; SM00091; PAS; 2.
DR TIGRFAMs; TIGR00229; sensory_box; 2.
DR PROSITE; PS00036; HLH_1; 1.
DR PROSITE; PS00888; HLH_2; 1.
DR PROSITE; PS0112; PAS; 2.
DR Repeat; DNA-binding; Nuclear protein; Transcription regulation;
KW Activator; Angiogenesis; Developmental protein.
FT DNA_BIND 15 27
FT DOMAIN 28 68
FT DOMAIN 84 154
FT DOMAIN 230 300
FT DOMAIN 304 347
FT DOMAIN 474 480
FT CONFLICT 60 60
FT CONFLICT 539 539
FT CONFLICT 601 601
FT CONFLICT 693 693
FT CONFLICT 716 716
FT CONFLICT 722 722
FT CONFLICT 765 765
FT CONFLICT 769 769
FT CONFLICT 844 844
FT CONFLICT 847 847
SQ SEQUENCE 870 AA; 96425 MW; 3DF5B7B13AEC871D CRC64;
Alignment Scores:
Pred. No.: 6.28e-33 Length: 870
Score: 572.50 Matches: 129
Percent Similarity: 64.34% Conservative: 37
Best Local Similarity: 50.00% Mismatches: 58
Query Match: 28.48% Indels: 34
DB: 1 Gaps: 7
```

```
US-09-896-791b-2 (1-1100) x PAS1_HUMAN (1-870)
QY 19 ATGGCGTGGGGTGCAGCGCGCTGAGTGCAGACACCGAGCTGCGGAAGGAGAGTGC GCG 78
   |||  :::::  |||||:::  |||  |||||:::  |||  |||||:::  |||  |||||:::
Db 1 MetThrAlaAspLysGluLysArgSerSerSerGluArgArgLysLysSerArg 20
QY 79 GAGCGCGCGCGAGCGCGCGAGCCAGGAGGAGGAGTCTGTACAGCTGCGCGCACACT 138
   |||  :::::  |||||:::  |||||:::  |||||:::  |||||:::  |||||:::  |||||:::
Db 21 AspAlaAlaArgCysArgSerLysGluThrGluValPheTyrGluLeuAlaHisGlu 40
QY 139 CTGCGCTTTGCGCGCGCGTACGCGCGCAGCCCTGGAGAGGCGCTCCATCATGCGCTCAC 198
   |||||:::  |||||:::  |||||:::  |||||:::  |||||:::  |||||:::  |||||:::  |||||:::
Db 41 LeuProLeuProHisSerValSerHisLeuAspLysAlaSerIleMetArgLeuAla 60
QY 199 ATCAGCTACTCGGCGATGACCGCGCTC-----TGGCGACAGCGTGGAAGAGG 246
   |||||:::  |||||:::  |||||:::  |||||:::  |||||:::  |||||:::  |||||:::  |||||:::
Db 61 IleSerPheLeuArgThrHisLysLeuLeuSerSerValCysSer-GluAsnGluSerG 80
QY 247 GGGAGAG-----CCACTGGACCGCTGCTACCTGAAGCGCCCTGGAGGTTTCGTCAT 297
   |||  :::::  |||||:::  |||||:::  |||||:::  |||||:::  |||||:::  |||||:::
Db 80 uAlaGluAlaAspGlnGlnMetAspAsnLeuTyrLeuLysAlaLeuGluGlyPheIleAl 100
QY 298 GGTACTCACCGCGAGGAGACATGCTTACCTGTCGGAAATGCTCAGCAAGCACCTGG 357
   |||||:::  |||||:::  |||||:::  |||||:::  |||||:::  |||||:::  |||||:::  |||||:::
Db 100 aValValThrGlnAspGlyAspMetIlePheLeuSerGluAsnIleSerLysPheMetG 120
QY 358 CCTCAGTCACTGGACCTCTGTTCTCCTCCTCCCTGATACATAACCCCACTCCTCGTACCAAT 417
   |||||:::  |||||:::  |||||:::  |||||:::  |||||:::  |||||:::  |||||:::  |||||:::
Db 120 yLeuThrGln----- 123
QY 418 TTCCTCTCTGGAGCTATTGGACACAGTATCTTTGATTTCATCCATCCCTGTGACCAAGAG 477
   |||  :::::  |||||:::  |||||:::  |||||:::  |||||:::  |||||:::  |||||:::
Db 124 ---Val--GluLeuThrGlyHisSerIlePheAspPheThrHisProcysAspHisGlu 141
QY 478 GRACITTCAGACGCCCTGACCCCGAG-----CCGAACCTGTCGAAGAGAGAGCTGGAA 531
   |||||:::  |||||:::  |||||:::  |||||:::  |||||:::  |||||:::  |||||:::  |||||:::
Db 142 GluIleArgGluAsnLeuSerLeuLysAsnGlySerGlyPheGlyLysSerLysAsp 161
QY 532 GCCCCCAACAGAGCGCCACATTTTCCCTGCGAATGAAGACAGCGCTCACACAGAGGCGCG 591
   |||||:::  |||||:::  |||||:::  |||||:::  |||||:::  |||||:::  |||||:::  |||||:::
Db 162 MetSerThrGluArgAspPhePheMetArgMetLysCysThrValThrAsnArgGlyArg 181
QY 592 ACCTCAACTCAAGCGCGCCACTGGAAGGTGCTGCATGCTCAGGACATATGAGGCGC 651
   |||||:::  |||||:::  |||||:::  |||||:::  |||||:::  |||||:::  |||||:::  |||||:::
Db 182 ThrValAsnLeuLysSerAlaThrTrpLysValLeuHisCysThrGlyGlnValLysVal 201
QY 652 TACAAGCCC---CCTGCACAGACTTCCCTCGCGGAGCGCTCCGCTCCGAGCTCCCTCG 708
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 202 TyrAsnAsnCysProHisAsnSerLeuCysGly-----TyrLysGluProLeuLeu 219
QY 709 CAATGCTGTGCTTATCTGTGAAGCCATCCCGACGCTCCCTCCCTCCAC 756
   |||||:::  |||||:::  |||||:::  |||||:::  |||||:::  |||||:::  |||||:::  |||||:::
Db 220 SerCysLeuIleIleMetCysGluProIle---GlnHisProSerHis 234
RESULT 5
SIMA_DROME
ID SIMA_DROME STANDARD; PRT; 1507 AA.
AC Q24167; O9VAA5;
DT 15-DEC-1998 (Rel. 37, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Similar protein.
GN SIMA OR CG7951.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;
OC Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
OC Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxId=7227;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=96269413; PubMed=8682312;
RA Nambu J.R., Chen W., Hu S., Crews S.T.;
RA "The Drosophila melanogaster similar bHLH-PAS gene encodes a protein
```

```
RT related to human hypoxia-inducible factor 1 alpha and Drosophila
RT single-minded.";
RT Gene 172:249-254 (1996).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Berkeley;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celnikier S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Balley R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotter P.,
RA Burtis K.C., Busan D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattel B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacle J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirska R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT The genome sequence of Drosophila melanogaster.";
RL Science 287:2185-2195 (2000).
CC -!- FUNCTION: POSSIBLE DNA-BINDING TRANSCRIPTIONAL ACTIVATOR.
CC -!- SUBUNIT: EFFICIENT DNA BINDING REQUIRES DIMERIZATION WITH ANOTHER
CC BHLH PROTEIN.
CC -!- SUBCELLULAR LOCATION: Nuclear (Potential).
CC -!- TISSUE SPECIFICITY: UBQUITOUSLY EXPRESSED IN THE EMBRYO.
CC -!- SIMILARITY: BELONGS TO THE BASIC HELIX-LOOP-HELIX (BHLH) FAMILY OF
CC TRANSCRIPTION FACTORS. HIGHEST TO HIF-1 ALPHA.
CC -!- SIMILARITY: CONTAINS 2 PAS (PER-ARNT-SIM) DIMERIZATION DOMAINS.
CC -!- SIMILARITY: CONTAINS 1 PAS-ASSOCIATED C-TERMINAL (PAC) DOMAIN.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
CC EMBL; U43090; AAC47303.1; -.
DR EMBL; AE003772; AAF57008.2; -.
DR FlyBase; FBgn0015542; sima.
DR InterPro; IPR001092; HLH_basic.
DR InterPro; IPR001610; PAC.
DR InterPro; IPR000014; PAS_domain.
DR Pfam; PF00785; PAC; 1.
DR Pfam; PF00989; PAS; 2.
DR SMART; SM00353; HLH; 1.
DR SMART; SM00086; PAC; 1.
```











```
Db 100 sileMetryrileSerGluThrAlaSerValHisLeuGlyLeuSerGln----- 116
Qy 379 TCCTCTCCCTGATACATACCCCACTCCTGGTACCAATTTCTCTGGAGCTCATTTGA 438
Db 117 -----Val--GluLeuThrGly 121
Qy 439 CACAGTATCTTTGATTTCATCCCTGTGACCAAGAGGAACTTCAAGAGCCCTGACC 498
Db 122 AsnSerIleThrGluThrIleHisProSerAspHisAspGluMetThrAlaValLeuThr 141
Qy 499 CCCAGGCGAACCCTCTCAAGAAGAGCTG---GAAGCCCCCAACAGAGCCCACTTTTCC 555
Db 142 AlaHisProProLeuHisHisLeuLeuGlnGluThrGluLeuGluArgSerPhePhe 161
Qy 556 CTGCAATGAAGAGACACCTCACAGCAGAGAGGGCCGCTCAACCTCAAGCGGCCAC 615
Db 162 LeuArgMetLysCysValLeuAlaLys-----ArgAsnAlaGlyLeuThrCysSerGly 179
Qy 616 TGAAGGCTGCTGCTCAGGACATATGAGGGCTACAAGCCCTCCGACAGACTTCC 675
Db 180 TyrLysValIleHisCysSerGlyThrLeuLysIleArgGlnTyrMetLeuAspMetSer 199
Qy 676 CTGCGCGGAGC-----GCTACTCTGGTCTT-----CCTCGCTCCGAGCTCCC 705
Db 200 LeuTyrAspSerCysTyrGlnIleValGlyLeuValAlaValGlyGlnSerLeuProPro 219
Qy 706 CTGCAATCCTGGTCTTATCTGTGAAGCCATCCCGAGCTCCCTTCCAGCATGTT--- 762
Db 220 -----SerAlaIleThrGluIleLysLeuHisSerAsnMet 231
Qy 763 -----GCTACTCTGGTCTT----- 777
Db 232 PheMetPheArgAlaSerLeuAspLeuLysLeuIlePheLeuAspSerArgValThrGlu 251
Qy 778 -----CCACAG-----GAGAAGACTCCCATCTCTACCTTA----- 807
Db 252 LeuThrGlyTyrGluProGlnAspLeuIleGluLysThrLeuTyrHisHisValHisGly 271
Qy 808 -----TTCACCCCTTTTGGAGGCACTACTTCTGTGTCAGAGGTGGCTGTGT 858
Db 272 CysAspThrPheHisLeuArgTyrAlaHisHisLeuLeuValLys----- 287
Qy 859 CAGGTGCTACAGGGAAAGGACTGAA-----TCCTCTCTCCCTCCA 900
Db 288 -----GlyGlnValThrThrLysTyrTyrArgLeuLeuSerLysLeuGlyGly 303
Qy 901 TGGGTGTGTGGGCC-----CTTAACCGGAAA 927
Db 304 TrpVal---TrpValGlnSerTyrAlaThrValValHisAsnSerArgSerArgPro 322
Qy 928 AATGTCTCTGCTAGGAGGAGTGAAGACATG----- 960
Db 323 HisCys-----IleValSerValAsnTyrValLeuThrAspValGluTyr 337
Qy 961 -----GCCAGCTATCTTAGCCAGAAACCCACAAATGTCTCCAAACACCAATAAG 1014
Db 338 LysGluLeuGlnLeuSerLeuAspGlnValSerThrLysSerGlnGluSerTrpArg 357
Qy 1015 ACCTCTCCTTGTAGGCACAGAGAAAAA----- 1074
Db 358 ThrThrLeuSerThrSerGlnGluThrArgLysSerAlaLysProLysAsnThrLysMet 377
Qy 1075 AAAAAAAA 1083
Db 378 LysThrLys 380

RESULT 9
SIM1_HUMAN
ID SIM1_HUMAN STANDARD; PRT; 766 AA.
AC P81133;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
```

```
15-JUN-2002 (Rel. 41, Last annotation update)
Single-minded homolog 1.
SIM1.
Homo sapiens (Human).
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
NCBI_TaxID=9606;
[1]
SEQUENCE FROM N.A.
MEDLINE=97343329; PubMed=9199934;
Chraist R., Scott H.S., Chen H., Kudoh J., Rossier C.,
Minoshima S., Wang Y., Shimizu N., Antonarakis S.E.;
"Cloning of two human homologs of the Drosophila single-minded gene
SIM1 on chromosome 6q and SIM2 on 21q within the Down syndrome
chromosomal region";
Genome Res. 7:615-624(1997).
-!- FUNCTION: TRANSCRIPTIONAL FACTOR THAT MAY HAVE PLEIOTROPIC EFFECTS
DURING EMBRYOGENESIS AND IN THE ADULT.
-!- SUBUNIT: EFFICIENT DNA BINDING REQUIRES DIMERIZATION WITH ANOTHER
BHLH PROTEIN. HETERODIMER OF SIM1 AND ARNT.
-!- SUBCELLULAR LOCATION: Nuclear (Potential).
-!- SIMILARITY: BELONGS TO THE BASIC HELIX-LOOP-HELIX (BHLH) FAMILY OF
TRANSCRIPTION FACTORS.
-!- SIMILARITY: CONTAINS 2 PAS (PER-ARNT-SIM) DIMERIZATION DOMAINS.
-!- SIMILARITY: CONTAINS 1 PAS-ASSOCIATED C-TERMINAL (PAC) DOMAIN.
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL Outstation
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----
EMBL; U070212; AAB62395.1; -.
TRANSFAC; T03729; -.
Genew; HGNC:10882; SIM1.
MIM; 603128; -.
DR InterPro; IPR001092; HLH_basic.
DR InterPro; IPR001610; PAC_domain.
DR InterPro; IPR000014; PAS_domain.
DR Pfam; PF00785; PAC; 1.
DR Pfam; PF00989; PAS; 2.
DR SMART; SM00353; HLH; 1.
DR SMART; SM00086; PAC; 1.
DR SMART; SM00091; PAS; 2.
DR PROSITE; PS00038; HLH_1; 1.
DR PROSITE; PS50112; PAS; 2.
DR PROSITE; PS50888; HLH_2; 1.
KW Developmental protein; Neurogenesis; Nuclear protein; Repeat;
Transcription regulation; DNA-binding.
FT DNA_BIND 1 13 BASIC DOMAIN.
FT DOMAIN 14 54 HELIX-LOOP-HELIX MOTIF (BY SIMILARITY).
FT DOMAIN 77 147 PAS 1.
FT DOMAIN 218 288 PAS 2.
SQ SEQUENCE 766 AA; 85474 MW; 6033694BF9A29DD6 CRC64;

Alignment Scores: 1.12e-14 Length: 766
Pred. No.: 312.50 Matches: 82
Score: 51.80% Conservative: 33
Percent Similarity: 36.94% Mismatches: 69
Best Local Similarity: 15.55% Indels: 38
Query Match: 1 Gaps: 4
DB:

US-09-896-791b-2 (1-1100) x SIM1_HUMAN (1-766)
Qy 64 AAGGAGAGTCGCGGAGCGCGCCGCGAGCGCGGAGGAGTGCTGTAC 123
Db 2 LysGluLysSerLysAsnAlaAlaArgThrArgGluLysGluAsnSerGluPheTyr 21
Qy 124 CAGCTGGCGCACACTCTGCCCTTTGCGCGCGCGCTGACGCGCACCTGACAAAGCCTCC 183
Db 124 CAGCTGGCGCACACTCTGCCCTTTGCGCGCGCGCTGACGCGCACCTGACAAAGCCTCC 183
```

```
Db 22 GluLeuAlaLysLeuLeuProLeuAlaSerAlaIleThrSerGlnValAspLysAlaSer 41
QY 184 ATCATCGCGCTCACATACAGTACTGCGCATGCACCGCGCTGCGCAGCAGGT----- 237
   |||:||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 42 IleIleArgLeuThrThrSerTyrLeuLysMetArgValValPheProGluGlyLeuGly 61
QY 238 -----GGAAAAAGGGGAGAGCCACTGGA 261
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 62 GluAlaTrpGlyHisSerSerArgThrSerProLeuAspAsnValGlyArgGluLeuGly 81
QY 262 CGCCTCTACCTCAAGCGCCTGGAGGGTTGCTCATGGTACTCACCGCCGAGGGAGACAT 321
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 82 SerHisLeu-LeuGlnThrLeuAspGlyPheIlePheValAlaProAspGlyLysI1 101
QY 322 GCCTTACCTGTGCGAAATGTCAGAACGACCTGGCGCTCAGTCAGTCAGCTGCGACCTCTGTTCC 381
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 101 eMetTyrIleSerGluThrAlaSerValHisLeuGlyLeuSerGln----- 116
QY 382 TCCTCCCTGATACATAACCCCACTCCTCGTGACCAATTTCTCTCGGAGCTCATTTGGACAC 441
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 117 -----Val--GluLeuThrGlyAsn 122
QY 442 AGTATCTTTGATTTATCATCCCTGTCAGCAAGAGGAACTTCAAGACGCCCTGACCCCC 501
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 123 SerIleTyrGluTyrIleHisProAlaAspHisAspGluMetThrAlaValLeuThrAla 142
QY 502 ---AGCCGAACCTGTGCAAGAAGAGCTGGAAGCGCCCAAGAGCGCCACTTTTCCCTG 558
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 143 HisGlnProTyrHisSerHisPheValGlnGluTyrGluIleGluArgSerPhePheLeu 162
QY 559 CGAATGAAGAGCAGCTCACCAGCAGAGGGCGCGCACCTCAACCTCAAGCGGCCACTGG 618
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 163 ArgMetLysCysValLeuAlaLys-----ArgAsnAlaGlyLeuThrCysGlyGlyTyr 180
QY 619 AGGTGCTGCATGCTCAGCATATGAGGGCTACAGCCCTCAGCAGACTTCCCT 678
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 181 LysValIleHisCysSerGlyTyrLeuLysIleArgGlnTyrSerLeuAspMetSerPro 200
RESULT 10
SIM2_HUMAN
ID SIM2_HUMAN STANDARD: PRT: 667 AA.
AC Q14190: Q15470: Q15471: Q15472: Q15473: Q16532;
DT 01-NOV-1997 (Rel. 35, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DE 15-JUN-2002 (Rel. 41, Last annotation update)
DE Single-minded homolog 2.
GN SIM2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97343329; PubMed=9199934;
RA Chast R., Scott H.S., Chen H., Kudoh J., Rossier C., Minoshima S.,
RA Wang Y., Shimizu N., Antonarakis S.E.;
RT "Cloning of two human homologs of the Drosophila single-minded gene
RT SIM1 on chromosome 6q and SIM2 on 21q within the Down syndrome
RT chromosomal region.";
RL Genome Res. 7:615-624(1997).
RN [2]
RP SEQUENCE FROM N.A.
RA Shibuya K., Kudoh J., Minoshima S., Kawasaki K., Nakatoh E.,
RA Shintani A., Asakawa S., Shimizu N.;
RT "Genomic sequencing of 1.2-Mb region on human chromosome 21q22.2.";
RL Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE=20289799; PubMed=10830953;
RA Hattori M., Fujiyama A., Taylor T.D., Watanabe H., Yada T.,
RA Park H.-S., Toyoda A., Ishii K., Totoki Y., Choi D.-K., Soeda E.,
RA Ohki M., Takagi T., Sakaki Y., Taudien S., Blechschmidt K., Polley A.,
RA Menzel U., Delabar J., Kumpf K., Lehmann R., Patterson D.,
```

```
RA Reichwald K., Rump A., Schillhabel M., Schudy A., Zimmermann W.,
RA Rosenthal A., Kudoh J., Shibuya K., Kawasaki K., Asakawa S.,
RA Shintani A., Sasaki T., Nagamine K., Mitsuyama S., Antonarakis S.E.,
RA Minoshima S., Shimizu N., Nordsiek G., Hornischer K., Brandt P.,
RA Scharfe M., Schoen O., Desario A., Reichelt J., Kauer G., Bloeker H.,
RA Ramser J., Beck A., Klages S., Hennig S., Riesselmann L., Dagand E.,
RA Wehmeyer S., Borzym K., Gardiner K., Nizetic D., Francis F.,
RA Lehrach H., Reinhardt R., Yaspo M.-L.;
RT "The DNA sequence of human chromosome 21.";
RL Nature 405:311-319(2000).
RN [4]
RP SEQUENCE OF 1-247 FROM N.A.
RA Osoegawa K., Okano S., Soeda E.;
RL Submitted (JUL-1996) to the EMBL/GenBank/DBJ databases.
RN [5]
RP SEQUENCE OF 1-58; 87-152; 183-247 AND 249-283 FROM N.A.
RX MEDLINE=9357794; PubMed=7647800;
RA Chen H., Chast R., Rossier C., Gos A., Antonarakis S.E., Kudoh J.,
RA Yamaki A., Shindoh N., Maeda H., Minoshima S., Shimizu N.;
RT "Single-minded and Down syndrome?";
RL Nat. Genet. 10:9-10(1995).
RN [6]
RP SEQUENCE OF 87-116 FROM N.A.
RX MEDLINE=96016135; PubMed=7568099;
RA Dahmane N., Chartron G., Lopes C., Yaspo M.-L., Maunoury C.,
RA Decorte L., Sinet P.M., Bloch B., Delabar J.M.;
RT "Down syndrome-critical region contains a gene homologous to
RT Drosophila sim expressed during rat and human central nervous system
RT development.";
RL Proc. Natl. Acad. Sci. U.S.A. 92:9191-9195(1995).
RN [7]
RP SEQUENCE OF 154-181 FROM N.A.
RX MEDLINE=96299750; PubMed=8661114;
RA Yamaki A., Noda S., Kudoh J., Shindoh N., Maeda H., Minoshima S.,
RA Kawasaki K., Shimizu Y., Shimizu N.;
RT "The mammalian single-minded (SIM) gene: mouse cDNA structure and
RT diencephalic expression indicate a candidate gene for Down
RT syndrome.";
RL Genomics 35:136-143(1996).
CC -!- FUNCTION: TRANSCRIPTION FACTOR THAT MAY BE A MASTER GENE OF CNS
CC DEVELOPMENT IN COOPERATION WITH ARNT. IT MAY HAVE PLEIOTROPIC
CC EFFECTS IN THE TISSUES EXPRESSED DURING DEVELOPMENT.
CC -!- SUBUNIT: EFFICIENT DNA BINDING REQUIRES DIMERIZATION WITH ANOTHER
CC BHLH PROTEIN. HETERODIMER OF SIM2 AND ARNT.
CC -!- SUBCELLULAR LOCATION: Nuclear (Potential).
CC -!- ALTERNATIVE PRODUCTS: 2 ISOFORMS; SIM2 (SHOWN HERE) AND SIM2S; ARE
CC PRODUCED BY ALTERNATIVE SPLICING.
CC -!- SIMILARITY: BELONGS TO THE BASIC HELIX-LOOP-HELIX (BHLH) FAMILY OF
CC TRANSCRIPTION FACTORS.
CC -!- SIMILARITY: CONTAINS 2 PAS (PER-ARNT-SIM) DIMERIZATION DOMAINS.
CC -!- SIMILARITY: CONTAINS 1 PAS-ASSOCIATED C-TERMINAL (PAC) DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U80456; AAB62396.1; -
DR EMBL; U80457; AAB62397.1; -
DR EMBL; AB003185; BAA21489.1; -
DR EMBL; AB003185; BAA21490.1; -
DR EMBL; D85922; BAA12919.1; -
DR EMBL; D44444; BAA07906.1; -
DR EMBL; D44445; BAA07907.1; -
DR EMBL; D44446; BAA07908.1; -
DR EMBL; D44447; BAA07909.1; -
DR EMBL; D44448; BAA07910.1; -
DR EMBL; X84790; CA559261.1; -
DR EMBL; D70838; BAA11081.1; -
DR EMBL; AP000697; BAA89433.1; -
```

DR EMBL; AP001126; -, NOT\_ANNOTATED\_CDS.  
 DR TRANSFAC; T04910; -  
 DR Genew; HGNC:10883; SIM2.  
 DR MIM; 600892; -  
 DR InterPro; IPR001092; HLH\_basic.  
 DR InterPro; IPR001610; PAC.  
 DR InterPro; IPR000014; PAS\_domain.  
 DR Pfam; PF00785; PAC; 1.  
 DR Pfam; PF00989; PAS; 1.  
 DR SMART; SM00353; HLH; 1.  
 DR SMART; SM00086; PAC; 1.  
 DR SMART; SM00091; PAS; 2.  
 DR PROSITE; PS00038; HLH\_1; 1.  
 DR PROSITE; PS00888; HLH\_2; 1.  
 DR PROSITE; PS0112; PAS; 2.  
 KW Developmental protein; Neurogenesis; Nuclear protein; Repeat;  
 KW Transcription regulation; DNA-binding; Alternative splicing.  
 FT DNA\_BIND 1 13  
 FT DOMAIN 14 54  
 FT DOMAIN 77 149  
 FT DOMAIN 218 288  
 FT VARSPLIC 453 667.  
 FT  
 FT ANPLVPSSSSPAKNPPEPPANTARHSLSVPSYEAAPAAVRRF  
 FT GEDTAPPSPSCGHYREEPALGPAPKAQAQAARDGARLALAR  
 FT AAPECCAPPTPEAPGAPALQPFVLLNYHRLARRGPIGAA  
 FT PAASGLACAPGPEAATGALRLRHPSPAATSPGAPLPHYL  
 FT GASVLIITNGR -> HEFSKKPMLPAKFGQPOGSPCEVARE  
 FT FLSTMPASGECOWHANYANPLVPSSSSPAKNPPEPPANTARHS  
 FT LVPSYEGSGLLVGVGLRTAGSRSSHGGMQWOMETEPSRF  
 FT GOTCPLSASK (IN ISOFORM SIM2S).  
 FT IH -> RI (IN REF. 5).  
 FT CONFLICT 183 184  
 FT SEQUENCE 667 AA; 73218 MW; A100880541A74E6B CRC64;  
 SQ  
 Alignment Scores:  
 Pred. No.: 1.43e-14 Length: 667  
 Score: 311.00 Matches: 124  
 Percent Similarity: 42.38% Conservative: 54  
 Best Local Similarity: 29.52% Mismatches: 123  
 Query Match: 15.47% Indels: 119  
 DB: 1 Gaps: 15  
 US-09-896-791b-2 (1-1100) x SIM2\_HUMAN (1-667)  
 QY 64 AAGGAGAAGTCGGGACGCGCGCCGACGCGGACGAGGAGGTGCTGTAC 123  
 Db 2 LysGluLysSerLysAsnAlaLysThrArgGluLysGluAsnGlyGluPheTyr 21  
 QY 124 CAGCTGGCCACACTGCTGCTTTCGCGCGCGCTGAGCGGACCTGGACAGGCTCC 183  
 Db 22 GluLeuAlaLysLeuLeuProLeuProSerAlaIleThrSerGlnLeuAspLysAlaSer 41  
 QY 184 ATCATGCGCTCACANTCAGTACCTGCGGATGCGACCGCTCTGCGCAGCAGGT----- 237  
 DQ 42 IleileArgLeuThrThrSerTyrLeuLysMetArgAlaValPheProGluGlyLeuGly 61  
 QY 238 -----GGAAAGGGGGAGAGCC-----ACTGGA 261  
 Db 62 AspAlaTrpGlyGlnProSerArgAlaGlyProLeuAspGlyValAlaLysGluLeuGly 81  
 QY 262 CGCCTGCTACCTGAAGCCCTGGAGGTTTCGTATGTTACTACCGCGGAGGAGACAT 321  
 Db 82 SerHisLeu-LeuGlnThrLeuAspGlyPheValPheValValAlaSerAspGlyLysIle 101  
 QY 322 GGTTCACCTGTCGGAANAATGTCAGCAAGCACCTGGCGCTCAGTCACTGGACCTCTCTCC 381  
 Db 101 eMetTyrIleSerGluThrAlaSerValHisLeuGlyLeuSerGln----- 116  
 QY 382 TCCTCCCTGTATACATAACCCACTCTCTGTACCAATTTCTCTGCGAGCTCATTCGACAC 441  
 Db 117 -----Val--GluLeuThrGlyAsn 122  
 QY 442 AGTATCTTTGATTATTCATCCTGTGTACCAAGAGGAACTTCAAGACGCGCTGACCCCC 501  
 Db 123 SerIleTyrGluTyrIleHisProSerAspHisAspGluMetThrAlaValLeuThrAla 142  
 QY 502 AGGCCGAACCTGTCAAGAAGAAGCTG---GAAGCCCCAACAGAGCGCCACATTTTCCCTG 558  
 Db 143 HisGlnProLeuHisHisLeuLeuGlnGluTyrGluIleGluArgSerPhePheLeu 162  
 QY 559 CGAATGAAGACGACGCTCACAGCAGAGCGGCGCACCTCAACCTCAAGCGCACCTGG 618  
 Db 163 ArgMetLysCysValLeuAlaLys-----ArgAsnAlaGlyLeuThrCysSerGlyTyr 180  
 QY 619 AAGGTGCTGACTGCTCAGACATATGAGGCGCTACAAGCCCTGCACAGACTTCCCT 678  
 Db 181 LysValIleHisCysSerGlyTyrLeuLysIleArgGlnTyrMetLeuAspMetSerLeu 200  
 QY 679 GCCGGGAGC-----CCTCGCTCCGAGCTCCCTG 708  
 Db 201 TyrAspSerCysTyrGlnIleValGlyLeuValAlaValGlyGlnSerLeuProPro--- 219  
 QY 709 CAATGCTGCTGCTTATCTGTGAAGCCATCCCGACCTCCCTTCCACGATGGT----- 762  
 Db 220 -----SerAlaIleThrGluIleLysLeuTyrSerAsnMetPhe 232  
 QY 763 -----GCTACTCTGGTCTT----- 777  
 Db 233 MetPheArgAlaSerLeuAspLeuLysLeuIlePheLeuAspSerArgValThrGluVal 252  
 QY 778 -----CCACAG-----GAGAAGACTCCCATC-----TCT 801  
 Db 253 ThrGlyTyrGluProGlnAspLeuIleGluLysThrLeuTyrHisHisValHisGlyCys 272  
 QY 802 ACCTTATTCACCCCTTTTGGAGGACACTACTTGTCTGTCAAGAGGTGGCGCTGT--- 858  
 Db 273 AspValPheHisLeuArgTyrAlaHisHisLeuLeuValLysGlyGlnValThrThr 292  
 QY 859 -----CAGTGTCTACAGGGGAAGG----- 879  
 Db 293 LysTyrTyrArgLeuLeuSerLysArgGlyGlyTyrValTyrValGlnSerTyrAlaThr 312  
 QY 880 -----ACTGAATCCTCTCTCCCT-----TCATGGGTGTG 909  
 Db 313 ValValHisAsnSerArgSerSerArgProHisCysIleValSerValAsnTyrValLeu 332  
 QY 910 TGGGCGCTTAACCGGAAATAATTGCTCTGCTAGGAGGAGTGAAGACATGGCCACAGTA 969  
 Db 333 ThrGluIleGluTyrLysGlu-----LeuGlnLeu 342  
 QY 970 TCCTTAGCCAGAACCCACAAATGTCTCAAAACACCATAAAGACCTCTCTCTGTAG 1029  
 Db 343 SerLeuGluGlnValSerThrAlaLysSerGlnAspSerTyrArgThrAlaLeuSerThr 362  
 QY 1030 GCACACAGAAAAA-----AAAAAAAAAAAAAAAAAAAAAAAAAAAA 1083  
 Db 363 SerGlnGluThrArgLysLeuValLysProLysAsnThrLysMetLysThrLys 380  
 RESULT 11  
 SIM1\_MOUSE STANDARD; PRT; 765 AA.  
 AC Q61045; P70183;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 15-JUL-1999 (Rel. 38, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Single-minded homolog 1 (mSIM1).  
 GN SIM1.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Swiss Webster;  
 RX MEDLINE=97020303; PubMed=8812055;  
 RA Fan C.-M., Kuwana E., Bulfone A., Fletcher C.F., Copeland N.G.,  
 RA Jenkins N.A., Crews S., Martinez S., Puellas L., Rubenstein J.L.,



Db 143 HisGlnProTyrHisSerHisPheValGlnGluTyrGluIleGluArgSerPhePheLeu 162  
 Oy CGAATGAAGACGACCTCACCAGCAGAGGGCGCGCTCAACCTCAAGCGGCCACCTGG 618  
 ||||| ||| ||| |||  
 Db 163 ArgMetLysCysValLeuAlaLys-----ArgAsnAlaGlyLeuThrCysGlyGlyTyr 180  
 |||||  
 Oy 619 AAGGTGCTGCACTGCTCAGGACATATGAGCGCCTACAGCCCTGCACAGACTTCCCTC 678  
 ||||| ||||| ||||| |||||  
 Db 181 LysValIleHisCysSerGlyTyrLeuLysIleArgGlnTyrSerLeuAspMetSerPro 200  
 |||||  
 RESULT 12  
 TRH\_DROME STANDARD; PRT; 958 AA.  
 AC Q24119; Q24165; Q9W0Q7;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Trachealeless protein.  
 GN TRH OR C66883.  
 OS Drosophila melanogaster (Fruit fly).  
 OC Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;  
 OC Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;  
 OC Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.  
 OX NCBI\_TaxID=7227;  
 [1]  
 RN SEQUENCE FROM N.A.  
 RP TISSUE=Embryo;  
 RC MEDLINE=96136712; PubMed=8557198;  
 RX Wilk R., Weizman I., Shilo B.-Z.;  
 RA "Trachealeless encodes a bHLH-PAS protein that is an inducer of  
 RT tracheal cell fates in Drosophila.";  
 RL Genes Dev. 10:93-102(1996).  
 [2]  
 RN SEQUENCE FROM N.A.  
 RP TISSUE=Embryo;  
 RC MEDLINE=96136713; PubMed=8557189;  
 RX Isaac D.D., Andrew D.J.;  
 RA "Tubulogenesis in Drosophila: a requirement for the trachealeless gene  
 RT product.";  
 RL Genes Dev. 10:103-117(1996).  
 [3]  
 RN SEQUENCE FROM N.A. (ISOFORM 1).  
 RP STRAIN=Berkley;  
 RC MEDLINE=20196006; PubMed=10731132;  
 RX Adams M.D., Celnikier S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
 RA Ananidis P.G., Scherer S.E., Richards S., Ashburner M., Henderson S.N.,  
 RA George R.A., Lewis S.E., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,  
 RA Sutton G.G., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,  
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,  
 RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkuch C., Baldwin D.,  
 RA Balow R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,  
 RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brothier P.,  
 RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,  
 RA Cherry J.M., Cavley S., Dahlke C., Davenport L.B., Davies P.,  
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,  
 RA Foster C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,  
 RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,  
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,  
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
 RA Klumel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
 RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,  
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,  
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
 RA Shue B.C., Siden-Klamis I., Simpson M., Skupski M.P., Smith T.,

RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
 RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
 RA Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,  
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Smith H.O.,  
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
 RT "The genome sequence of Drosophila melanogaster.";  
 RL Science 287:2185-2195(2000).  
 CC -1- FUNCTION: TRANSCRIPTION FACTOR, MASTER REGULATOR OF TRACHEAL CELL  
 CC FATES IN THE EMBRYO, NECESSARY FOR THE DEVELOPMENT OF THE SALIVARY  
 CC GLAND DUCT AND THE POSTERIOR SPIRACLES. IT MAY INDUCE A GENERAL  
 CC FATE OF BRANCHED TUBULAR STRUCTURES OF EPITHELIAL ORIGIN. TCO/TRH  
 CC HETERODIMERS ARE INVOLVED IN THE CONTROL OF BREATHLESS EXPRESSION.  
 CC -1- SUBUNIT: EFFICIENT DNA BINDING REQUIRES DIMERIZATION WITH ANOTHER  
 CC BHLH PROTEIN. HETERODIMER WITH TCO.  
 CC -1- SUBCELLULAR LOCATION: Nuclear.  
 CC -1- ALTERNATIVE PRODUCTS: AT LEAST 3 ISOFORMS; 1 (SHOWN HERE), 2 AND  
 CC 3; ARE PRODUCED BY ALTERNATIVE SPLICING.  
 CC -1- TISSUE SPECIFICITY: TRACHEA, SALIVARY GLAND DUCTS, POSTERIOR  
 CC SPIRACLES (FILZKOEPER PRIMORDIA) AND A SUBSET OF CELLS IN THE CNS.  
 CC -1- DEVELOPMENTAL STAGE: DURING EMBRYOGENESIS, FIRST DETECTED IN THE  
 CC TRACHEAL PLACODES AT STAGE 8, AND EXPRESSION CONTINUES THROUGHOUT  
 CC EMBRYONIC AND LARVAL DEVELOPMENT. IN THE DEVELOPING SALIVARY  
 CC GLAND, EXPRESSION IS OBSERVED IN THE ENTIRE GLAND AT STAGE 9 AND  
 CC BY STAGE 12, EXPRESSION IS CONFINED TO THE SALIVARY DUCTS.  
 CC -1- SIMILARITY: BELONGS TO THE BASIC HELIX-LOOP-HELIX (BHLH) FAMILY OF  
 CC TRANSCRIPTION FACTORS.  
 CC -1- SIMILARITY: CONTAINS 2 PAS (PER-ARNT-SIM) DIMERIZATION DOMAINS.  
 CC -1- SIMILARITY: CONTAINS 1 PAS-ASSOCIATED C-TERMINAL (PAC) DOMAIN.  
 CC -1- CAUTION: REF.2 SEQUENCE DIFFERS FROM THAT SHOWN DUE TO FRAMESHIFTS  
 CC IN POSITIONS 22 TO 34.  
 -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 -----  
 CC EMBL; U33427; AAA96257.1; ALT\_INIT.  
 CC EMBL; U42699; AAA96754.1; ALT\_FRAME.  
 CC EMBL; AE003468; AAF47386.1; -  
 CC FlyBase; FBgn0003749; trh.  
 CC InterPro; IPR001092; HLH\_basic.  
 CC InterPro; IPR001610; PAC.  
 CC InterPro; IPR000014; PAS\_domain.  
 CC Pfam; PF00989; PAS; 2.  
 CC SMART; SM00353; HLH; 1.  
 CC SMART; SM00086; PAC; 1.  
 CC SMART; SM00091; PAS; 2.  
 CC PROSITE; PS00038; HLH\_1; 1.  
 CC PROSITE; PS00888; HLH\_2; 1.  
 CC PROSITE; PS0112; PAS; 2.  
 CC Developmental protein; Nuclear protein; Transcription regulation;  
 KW Repeat; DNA-binding; Alternative splicing.  
 KW DNA\_BIND 86 99 BASIC DOMAIN.  
 FT DOMAIN 100 140 HELIX-LOOP-HELIX MOTIF (BY SIMILARITY).  
 FT DOMAIN 174 244 PAS 1.  
 FT DOMAIN 391 461 PAS 2.  
 FT DOMAIN 465 508 PAC.  
 FT DOMAIN 629 636 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).  
 FT DOMAIN 154 157 POLY-SER.  
 FT DOMAIN 251 254 POLY-GLY.  
 FT DOMAIN 255 264 POLY-GLY.  
 FT DOMAIN 722 728 POLY-GLN.  
 FT VARSPIC 281 286 MISSING (IN ISOFORM 2).  
 FT VARSPIC 328 356 MISSING (IN ISOFORM 3).  
 FT CONFLICT 78 78 P -> A (IN REF. 1).  
 FT CONFLICT 250 250 G -> GG (IN REF. 2).  
 FT CONFLICT 703 703 A -> T (IN REF. 1).









RT "Aryl-hydrocarbon receptor nuclear translocator 2 (ARNT2): structure,  
 RL gene mapping, and candidate evaluation for human orofacial clefts."; ;  
 RN Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.  
 RP [2]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Brain;  
 RA MEDLINE=97349984; PubMed=9205841;  
 RA Nagase T., Ishikawa K.-I., Nakajima D., Ohira M., Seki N.,  
 RA Miyajima N., Tanaka A., Kotani H., Nomura N., Ohara O.;  
 RT "Prediction of the coding sequences of unidentified human genes. VII.  
 RT The complete sequences of 100 new cDNA clones from brain which can  
 RT code for large proteins in vitro."; ;  
 RL DNA Res. 4:141-150(1997).  
 CC -!- FUNCTION: SPECIFICALLY RECOGNIZES THE XENOBIOTIC RESPONSE ELEMENT  
 CC (XRE).  
 CC -!- SUBUNIT: EFFICIENT DNA BINDING REQUIRES DIMERIZATION WITH ANOTHER  
 CC BHLH PROTEIN. HETERODIMER WITH THE ARYL HYDROCARBON RECEPTOR (AHR)  
 CC OR THE SIM1 PROTEIN (By similarity).  
 CC -!- SUBCELLULAR LOCATION: Nuclear (Potential).  
 CC -!- SIMILARITY: BELONGS TO THE BASIC HELIX-LOOP-HELIX (BHLH) FAMILY OF  
 CC TRANSCRIPTION FACTORS. HIGHEST TO THE ARNT PROTEINS.  
 CC -!- SIMILARITY: CONTAINS 2 PAS (PER-ARNT-SIM) DIMERIZATION DOMAINS.  
 CC -!- SIMILARITY: CONTAINS 1 PAS-ASSOCIATED C-TERMINAL (PAC) DOMAIN.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; AF185610; AAG15310.1; -.  
 DR EMBL; AF185593; AAG15310.1; JOINED.  
 DR EMBL; AF185594; AAG15310.1; JOINED.  
 DR EMBL; AF185595; AAG15310.1; JOINED.  
 DR EMBL; AF185596; AAG15310.1; JOINED.  
 DR EMBL; AF185597; AAG15310.1; JOINED.  
 DR EMBL; AF185598; AAG15310.1; JOINED.  
 DR EMBL; AF185599; AAG15310.1; JOINED.  
 DR EMBL; AF185600; AAG15310.1; JOINED.  
 DR EMBL; AF185601; AAG15310.1; JOINED.  
 DR EMBL; AF185602; AAG15310.1; JOINED.  
 DR EMBL; AF185603; AAG15310.1; JOINED.  
 DR EMBL; AF185604; AAG15310.1; JOINED.  
 DR EMBL; AF185605; AAG15310.1; JOINED.  
 DR EMBL; AF185606; AAG15310.1; JOINED.  
 DR EMBL; AF185607; AAG15310.1; JOINED.  
 DR EMBL; AF185608; AAG15310.1; JOINED.  
 DR EMBL; AF185609; AAG15310.1; JOINED.  
 DR EMBL; AB002305; BAA20766.1; -.  
 DR TRANSFAC; T05059; -.  
 DR Genew; HGNC; 16876; ARNT2.  
 DR MIM; 606036; -.  
 DR InterPro; IPR001092; HLH\_basic.  
 DR InterPro; IPR001067; Nuc\_translocat.  
 DR InterPro; IPR001610; PAC\_translocat.  
 DR InterPro; IPR000014; PAS\_domain.  
 DR Pfam; PF00010; HLH; 1.  
 DR Pfam; PF00989; PAS; 2.  
 DR PRINTS; PR00785; NCTRNLOCATR.  
 DR SMART; SM00353; HLH; 1.  
 DR SMART; SM00086; PAC; 1.  
 DR SMART; SM00091; PAS; 2.  
 DR TIGRFAMs; TIGR00229; sensory\_box; 1.  
 DR PROSITE; PS00038; HLH\_1; 1.  
 DR PROSITE; PS00888; HLH\_2; 1.  
 DR PROSITE; PS0112; PAS; 2.  
 KW DNA-binding; Nuclear protein; Transcription regulation; Repeat.  
 FT DNA\_BIND 53 65  
 FT DOMAIN 66 106 BASIC DOMAIN.  
 FT DOMAIN 123 198 HELIX-LOOP-HELIX MOTIF (BY SIMILARITY).  
 FT PAS 1.  
 FT PAS 2.  
 FT DOMAIN 312 382

FT	DOMAIN	387	430	PAC.
FT	DOMAIN	62	65	POLY-ARG.
FT	DOMAIN	429	434	POLY-GLN.
FT	DOMAIN	541	544	POLY-SER.
FT	CONFLICT	64	64	R -> K (IN REF. 2).
FT	CONFLICT	668	668	G -> S (IN REF. 2).
SO	SEQUENCE	706 AA;	77612 MW;	2187A3C603EC7D39 CRC64;

Alignment Scores:  
 Pred. No.: 0.000254 Length: 706  
 Score: 165.00 Matches: 63  
 Percent Similarity: 42.86% Conservative: 57  
 Best Local Similarity: 22.50% Mismatches: 80  
 Query Match: 8.21% Indels: 81  
 DB: 1 Gaps: 10

US-09-896-791B-2 (1-1100) x ARNT2\_HUMAN (1-706)  
 Qy 52 ACCGAGCTCGGAAGGAGGAGTCCGGGAGCGCGCGCGCGCGCGCGCGCGAGGAGC 111  
 Db 50 SerLysPheSerArgGluAsnHisSerGluIleGluArgArgArgAsnLysMetThr 69  
 Qy 112 GAGTGTGTACCACTGCGGCACACTCTGCCTTTGGCGGGCGGTGAGCGCGCAGCTG 171  
 Db 70 GlnTyrIleThrGluLeuSerAspMetValProThrCysSerAlaLeuAlaArgLysPro 89  
 Qy 172 GACAGGCTCCATCATCGCTCACAATCAGTACCTACCTGCGCATCGCAGCGCTCTGCGCA 231  
 Db 90 AspLysLeuThrIleLeuArgMetAlaValSerHisMetLysSerMetArg----- 106  
 Qy 232 GCAGTGTGAAAAAGGGGAGGAGCACCCTGAGCGCTGCTAC----- 271  
 Db 107 -----GlyThrGlyAsnLysSerThr-AspGlyAlaTyrLysProSerPheLeuThrGl 124  
 Qy 272 -----CTGAGGCGCTGGAGGGTTTGTGATGGTACTCAGCGC 309  
 Db 124 uGlnGluLeuLysHisLeuIleLeuGluAlaAlaAspGlyPheLeuPheValAlaAla 144  
 Qy 310 CGAG--GGAGACATGGCTTACCTGTCGGAATCTCAGCAAGCAGCCTGGG---CCTCA 362  
 Db 144 aGluThrGlyArgValIleTyrValSerAspSerValThrProValLeuAsnGlnProGl 164  
 Qy 363 GTCAGTGGACCTCTGCTCTCTCTCCCTGATACATAACCCACCTCCTGGTACCAATTCTC 422  
 Db 164 nSer----- 165  
 Qy 423 TCTGGAGCTCATTTGACACAGTATCTTTGATTTTATCCATCCTGTGACCAAGAGAACT 482  
 Db 166 ---GluTrpPheGlySerThrLeuTyrGluGlnValHisProAspAspValGluLysLe 184  
 Qy 483 TCAAGACGCCCTG---ACCCCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 539  
 Db 184 uArgGluGlnLeuCysThrSerGluAsnSerMetThrGlyArgIleLeuAspLeuLysTh 204  
 Qy 540 A-----GAGCG 545  
 Db 204 rGlyThrValLysLysGluGlyGlnSerSerMetArgMetCysMetGlySerArgAr 224  
 Qy 546 CCACCTTTTCCCTGCGAATAAG----- 567  
 Db 224 gSerPheIleCysArgMetArgCysGlyAsnAlaProLeuAspHisLeuProLeuAsnAr 244  
 Qy 568 ----AGCAGCTCACCAGCAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 611  
 Db 244 gIleThrThrMetArgLysArgPheArgAsnGlyLeuGlyProValLysGluGlyGluAl 264  
 Qy 612 CACCTGGAAGTGTGTGCTGCTCAGGACATATGAGGCGCTACAGCCCGCTGCGCACAGC 671  
 Db 264 aGlnTyrAlaValHisCysThrGlyTyrIleLysAlaTrpProProAlaGlyMetH 284  
 Qy 672 TTCCTC---GCCGGGAGCGCTCGCTCGGAGCGCTCCCTGCAATGCGTGGTCTTATC 726  
 Db 284 rIleProGluGluAspAlaAspValGlyGlnGlySerLysTyrCysLeuValAlaIle 303

Search completed: December 2, 2002, 20:37:04  
Job time : 35 secs

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 Compugen Ltd.

OM nucleic - protein search, using frame\_plus\_n2p model

Run on: December 2, 2002, 20:29:55 ; Search time 61 Seconds  
(without alignments)  
7431.207 Million cell updates/sec

Title: US-09-896-791B-2

Perfect score: 2010

Sequence: 1 gaattcgacagggccat.....aaaaaaaaaacatgcggcgc 1100

Scoring table: BLOSUM62

Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 1343160

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+n2p model -DEV=xlp  
-Q=/cgn2.1/USPTO-SP001/US09896791/runat\_02122002\_152658\_5586/app\_query.fasta\_1.1287  
-DB=SPTREMBL\_21 -OFMT=fastan -SUFFIX=rspt -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0  
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45  
-DOALIGN=200 -THR\_SCORE=pct -THR\_MAX=100 -THR\_MIN=0 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US09896791 -ACGN\_1\_1.66 @runat\_02122002\_152658\_5586 -NCPU=6 -ICPU=3  
-NO\_XLPXY -NO\_MMAP -LARGQUERY -NEG\_SCORES=0 -WAIT -LONGLOG -DEV\_TIMEOUT=120  
-WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7  
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

SPTREMBL\_21.\*  
1: sp\_archaea.\*  
2: sp\_bacteria.\*  
3: sp\_fungi.\*  
4: sp\_humans.\*  
5: sp\_invertebrate.\*  
6: sp\_mammal.\*  
7: sp\_mhc.\*  
8: sp\_organelle.\*  
9: sp\_phage.\*  
10: sp\_plant.\*  
11: sp\_rodent.\*  
12: sp\_virus.\*  
13: sp\_vertebrate.\*  
14: sp\_unclassified.\*  
15: sp\_rvirus.\*  
16: sp\_bacteriap.\*  
17: sp\_archaeap.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Match	Length	ID	Description
1	1636	81.4	307	11 Q8VHR1	Q8VHR1 mus musculu

2	1045	52.0	630	11	Q9QX54	Q9qx54 mus musculu
3	1045	52.0	662	11	Q9Z2I5	Q9z2i5 mus musculu
4	1015.5	50.5	632	4	Q8WXA1	Q8wxal homo sapien
5	1011.5	50.3	632	4	Q96K34	Q96k34 homo sapien
6	1007	50.1	363	4	Q8TAP5	Q8tap5 homo sapien
7	1001	49.8	662	11	Q9JHS2	Q9jhs2 rattus norv
8	982.5	48.9	667	4	Q9Y2N7	Q9y2n7 homo sapien
9	753.5	37.5	648	4	Q9HAI2	Q9hai2 homo sapien
10	672	33.4	199	4	Q9HAM5	Q9ham5 homo sapien
11	613.5	30.5	258	11	Q9CYA8	Q9cya8 mus musculu
12	613.5	30.5	836	11	Q8R385	Q8r385 mus musculu
13	611.5	30.4	823	6	Q9ATA5	Q9ata5 bos taurus
14	607.5	30.2	811	13	Q9YIB9	Q9yib9 gallus gall
15	606.5	30.2	735	4	Q96PT9	Q96pt9 homo sapien
16	606.5	30.2	826	4	Q9UPB1	Q9upb1 homo sapien
17	603.5	30.0	823	11	Q9WTU9	Q9wtu9 rattus norv
18	593.5	29.5	825	11	Q35800	Q35800 rattus norv
19	587.5	29.2	874	11	Q9JHS1	Q9jhs1 rattus norv
20	575.5	28.6	867	13	Q9W7C6	Q9w7c6 gallus gall
21	575.5	28.6	870	13	Q9PTB3	Q9ptb3 coturnix co
22	572.5	28.5	870	6	Q9XTA4	Q9xta4 bos taurus
23	569.5	28.3	237	4	Q9H7Z9	Q9h7z9 homo sapien
24	565	28.1	873	13	Q8QGM4	Q8qgm4 fundulus he
25	547.5	27.2	766	13	Q98SW2	Q98sw2 oncorhynch
26	511.5	25.4	115	4	Q95262	Q95262 homo sapien
27	501.5	25.0	112	4	Q9UPH7	Q9uph7 homo sapien
28	424	21.1	805	13	Q918A9	Q918a9 xenopus lae
29	360	17.9	65	11	Q8RAD6	Q8rad6 mus musculu
30	343.5	17.1	590	4	Q9BY83	Q9by83 homo sapien
31	327	16.3	925	11	Q9QZQ0	Q9qzq0 mus musculu
32	321	16.0	657	11	Q35391	Q35391 mus musculu
33	319	15.9	901	4	Q9H323	Q9h323 homo sapien
34	318	15.8	903	4	Q9BY81	Q9by81 homo sapien
35	311	15.5	902	5	Q8SX13	Q8sx13 drosophila
36	309.5	15.4	585	13	Q8UVV3	Q8uvv3 brachydanio
37	308.5	15.3	765	11	Q70284	Q70284 mus musculu
38	300.5	15.0	760	13	Q9DDU6	Q9ddu6 xenopus lae
39	294.5	14.7	745	13	Q98SJ5	Q98sj5 brachydanio
40	293.5	14.6	849	5	O15984	O15984 bombyx mori
41	269.5	13.4	103	6	Q9N110	Q9n110 ovis aries
42	258.5	12.9	108	11	Q9QZ94	Q9qz94 cavia porce
43	216.5	10.8	719	5	Q963J8	Q963j8 caenorhabdi
44	216.5	10.8	823	5	O45486	O45486 caenorhabdi
45	216.5	10.8	825	5	Q9TVM0	Q9tvm0 caenorhabdi

#### ALIGNMENTS

RESULT 1  
Q8VHR1 PRELIMINARY; PRT; 307 AA.  
ID Q8VHR1  
AC Q8VHR1  
DT 01-MAR-2002 (TREMREL. 20, Created)  
DT 01-MAR-2002 (TREMREL. 20, Last sequence update)  
DT 01-JUN-2002 (TREMREL. 21, Last annotation update)  
DE Inhibitory PAS domain protein.  
GN IPAS  
OS Mus musculus (Mouse)  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=C57BL/6J;  
RX PubMed=11734856;  
RA Makino Y., Cao R., Svensson K., Bertilsson G., Asman M., Tanaka H.,  
Cao Y., Berkenstam A., Poellinger L.;  
RT "Inhibitory PAS domain protein is a negative regulator of hypoxia-  
inducible gene expression.";  
RL Nature 414:550-554(2001).  
DR EMBL; AF416641; AAL39015.1; -.  
DR InterPro; IPR001092; HLH\_basic.  
DR SMART; SM00353; HLH; 1.

DR PROSITE; PS00038; HELIX\_LOOP\_HELIX; UNKNOWN\_1.  
SQ SEQUENCE 307 AA; 33990 MW; D89D3B225C93967 CRC64;

## Alignment Scores:

Pred. No.: 3,61e-136 Length: 307  
Score: 1636.00 Matches: 307  
Percent Similarity: 100.00% Conservativity: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 81.39% Indels: 0  
DB: 11 Gaps: 0

US-09-896-791B-2 (1-1100) x Q8VHR1 (1-307)

QY 19 ATGCGGTGGGCTGAGCGCTGAGTGAACACCGAGTGGCGGAGGAGAGTGGCGG 78  
DB 1 MetAlaLeuGlyLeuGlnArgValArgSerAsnThrGluLeuArgLysGluLysSerArg 20  
QY 79 GACGCGCGCGGACGCGCGGACGCCAGGAGACGGAGTGGTGTACCAGCTGGCGGCACACT 138  
DB 21 AspAlaAlaArgSerArgSerGlnGluThrGluValLeuTyrGlnLeuAlaHisThr 40  
QY 139 CTGCGCTTTGCGCGCGGCTGACGCGCGCCTGAGCAAGGCTCCATCATCGCGCTCACA 198  
DB 41 LeuProPheAlaArgGlyValSerAlaHisLeuAspLysAlaSerIleMetArgLeuThr 60  
QY 199 ATCAGCTACCTGCGCATGACCCGCTCTGCGCAGCAGGTGGAAGGGGAGGAGGCACT 258  
DB 61 IleSerTyrLeuArgMetHisArgLeuCysAlaAlaGlyLysArgGlyArgAlaThr 80  
QY 259 GGACGCTGTACTACTGAAGCCCTGGAGGGTTTCATGTTACTCACCGCCGAGGAGAGA 318  
DB 81 GlyArgLeuLeuProGluGlyProGlyGlyPheArgHisGlyThrHisArgArgGlyArg 100  
QY 319 CATGGCTTACTCTCGGAATGTCAGCAGCAGCTGGGCTCAGTCAGTGGACCTCTGT 378  
DB 101 HisGlyLeuProValGlyLysCysGlnGlnAlaProGlyProGlnSerValAspLeuCys 120  
QY 379 TCCTCTCTCTGATACATAACCCCTCTCTGTTGTTACCAATTTCTCTGAGCTCATTTGA 438  
DB 121 SerSerSerLeuIleHisAsnProThrProGlyThrAsnPheSerLeuGluLeuIleGly 140  
QY 439 CACAGTATCTTTGATTTTATCCATTCCTGTGACCAAGAGGAATTCACAGCGCCCTGACC 498  
DB 141 HisSerIlePheAspPheIleHisProCysAspGlnGluLeuGlnAspAlaLeuThr 160  
QY 499 CCCAGCGGACCTGTCAAGAGAGAGCTGGAAGCCCCCAACAGCGCCACTTTTCCCTG 558  
DB 161 ProArgProAsnLeuSerLysLysLysLeuGluAlaProThrGluArgHisPheSerLeu 180  
QY 559 CGAATGAAGAGCAGCTCACAGCAGAGGCGGCACGCTCAACCTCAAAGCGGCGCACTGG 618  
DB 181 ArgMetLysSerThrLeuThrSerArgGlyArgThrLeuAsnLeuLysAlaAlaThrTrp 200  
QY 619 AAGTGCTGCATGCTCAGGACATATAGGGGCTTACAGGCCCTTACAGCCCTGCACAGACTTCCCT 678  
DB 201 LysValLeuHisCysSerGlyHisMetArgAlaTyrLysProAlaGlnThrSerPro 220  
QY 679 GCGGAGGCGCTCGTCCGAGCTCCCTGCAATGCTGCTGCTGCTGCTGCTGCTGCTGCTG 738  
DB 221 AlaGlySerProArgSerGluProProLeuGlnCysLeuValLeuIleCysGluAlaIle 240  
QY 739 CCCAGCTCCCTTCCACGATGCTGCTACTCTGCTGCTTCCACAGGAGAGACTCCCATC 798  
DB 241 ProGlnLeuProPheHisAspGlyAlaThrLeuGlyLeuProGlnGluLysThrProIle 260  
QY 799 TCTACCTTATACCTCTTTTGAAGGCACTACTTTGCTTGTCTCAAGAGTGGCTGTT 858  
DB 261 SerThrLeuPheThrProLeuTrpLysAlaLeuLeuCysLeuValLysArgTrpProVal 280  
QY 859 CAGGTGCTACAGGGGAAAGGACTGAATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 918  
DB 281 GlnValLeuGlnGlyLysGlyThrGluSerSerLeuProSerTrpValLeuTrpAlaLeu 300

QY 919 AACCGGAAAAATGTCTCTGGC 939  
DB 301 AsnArgLysAsnCysProGly 307  
RESULT 2  
Q90X54  
ID Q90X54 PRELIMINARY; PRT; 630 AA.  
AC Q90X54;  
DT 01-MAY-2000 (TREMBlrel. 13, Created)  
DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)  
DT 01-MAY-2000 (TREMBlrel. 13, Last annotation update)  
DE Hypoxia-inducible factor 3 alpha (Fragment).  
GN HIF3A.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
SEQUENCE FROM N.A.  
RA Gu Y.-Z., Moran S.M., Hogenesch J.B., Wartman L., Bradfield C.A.;  
RT "Cloning and Characterization of a Third Hypoxia Inducible Factor,  
HIF3-alpha.";  
RL J. Biol. Chem. 0:0-0(1999).  
DR EMBL; AF079153; AAF21782.1; JOINED.  
DR EMBL; AF079140; AAF21782.1; JOINED.  
DR EMBL; AF079141; AAF21782.1; JOINED.  
DR EMBL; AF079142; AAF21782.1; JOINED.  
DR EMBL; AF079143; AAF21782.1; JOINED.  
DR EMBL; AF079144; AAF21782.1; JOINED.  
DR EMBL; AF079145; AAF21782.1; JOINED.  
DR EMBL; AF079146; AAF21782.1; JOINED.  
DR EMBL; AF079147; AAF21782.1; JOINED.  
DR EMBL; AF079148; AAF21782.1; JOINED.  
DR EMBL; AF079149; AAF21782.1; JOINED.  
DR EMBL; AF079150; AAF21782.1; JOINED.  
DR EMBL; AF079151; AAF21782.1; JOINED.  
DR EMBL; AF079152; AAF21782.1; JOINED.  
DR MGD; MGI:1859778; Hif3a.  
DR InterPro; IPR001092; HLH\_basic.  
DR InterPro; IPR001067; Nuc\_translocat.  
DR InterPro; IPR001610; PAC.  
DR InterPro; IPR000014; PAS\_domain.  
DR Pfam; PF00989; PAS; 2.  
DR PRINTS; PR00785; NCTRNSLOCATR.  
DR SMART; SM00353; HLH; 1.  
DR SMART; SM00086; PAC; 1.  
DR SMART; SM00091; PAS; 2.  
DR PROSITE; PS00038; HELIX\_LOOP\_HELIX; UNKNOWN\_1.  
FT NON\_TER 630  
SQ SEQUENCE 630 AA; 69624 MW; 828EB2CB4E6D45B6 CRC64;

Alignment Scores:  
Pred. No.: 7,86e-84 Length: 630  
Score: 1045.00 Matches: 218  
Percent Similarity: 85.88% Conservativity: 1  
Best Local Similarity: 85.49% Mismatches: 5  
Query Match: 51.99% Indels: 31  
DB: 11 Gaps: 3

US-09-896-791B-2 (1-1100) x Q90X54 (1-630)

QY 43 AGTCTGAACACCGAGTGGCGGAGGAGAGTGGCGGAGCGCGCGCGCGCGCAGC 102  
DB 7 ArgSerAsnThrGluLeuArgLysGluLysSerArgAspAlaAlaArgSerArgSer 26  
QY 103 CAGGAGACGAGGTGTGTACACAGTGGCGGCACACTCTGCGCTTGGCGGCGGCGTACG 162  
DB 27 GlnGluThrGluValLeuTyrGlnLeuAlaHisThrLeuProPheAlaArgGlyValSer 46  
QY 163 GCGCACCTGCACAGCGCTCCATCATGCGCTCACATCAGCTACCTGCGCGCGCGCCG 222  
DB 47 AlaHisLeuAspLysAlaSerIleMetArgLeuThrIleSerTyrLeuArgMetHisArg 66

```
QY 223 CTCCTCCGACGAGTGGAA-----AAAGGGGAGAGCCACTGGAGCCCTG 267
Db 67 LeuCysAlaAlaGly-GluTrpAsnGlnValGluLysGlyGluProLeuAspAlaCy 86
QY 268 CTACCTGAAGCCCTGGAGGTTTCGTATGTTACTACCGCCGAGGAGACATGGCTTA 327
Db 86 sTyrLeuLysAlaLeuGluGlyPheValMetValLeuThrAlaGluGlyAspMetAlaTy 106
QY 328 CCTGTGCGAAATGTCAGACACCTGGGCTCAGTCAGTGACCTCTGTTCCCTCC 387
Db 106 rLeuSerGluAsnValSerLysHisLeuGlyLeuSerGln----- 119
QY 388 CTGATACATAACCCCACTCTGGTACCAATTTCTCTCGAGCTCATGGACACAGTATC 447
Db 120 -----Leu-GluLeuIleGlyHisSerIle 127
QY 448 TTTGATTTATCCATCCCTGTGACCAAGAGGAACTTCAAGAGCCCTGACCCCGAGCCG 507
Db 128 PheAspPheIleHisProCysAspGlnGluLeuGlnAspAlaLeuThrProArgPro 147
QY 508 AACCTGTCAAAGAAGAAGCTGGAAGCCCAACAGAGCGCCACTTTTCCCTCGGAATGAAG 567
Db 148 AsnLeuSerLysLysLeuGluAlaProThrGluArgHisPheSerLeuArgMetLys 167
QY 568 AGCAGCTCACACGAGAGGGCCCTACAGCCCAACAGCGGCCACTTCCCTGCGAGGTGCTG 627
Db 168 SerThrLeuThrSerArgGlyArgThrLeuAsnLeuLysAlaAlaThrTrpLysValLeu 187
QY 628 CACTGCTCAGGACATATGAGGGCCCTACAGCCCTGCAATGCTTATCTGTGAAGCCATCCC----- 741
Db 208 ProArgSerGluProProLeuGlnCysLeuValLeuIleCysGluAlaIleProHisPro 227
QY 742 -----CAGCTCCCTTCCACGATGGTGCTACTCTG 771
Db 228 AlaSerLeuGluProProLeuGlyArgGlyAlaPheLeu 240
RESULT 3
Q92215 PRELIMINARY; PRT; 662 AA.
AC Q92215;
DT 01-MAY-1999 (TReMBLrel. 10, Created)
DT 01-MAY-1999 (TReMBLrel. 10, Last sequence update)
DT 01-MAR-2002 (TReMBLrel. 20, Last annotation update)
DE Hypoxia inducible factor three alpha.
GN HIF3A.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN: [1]
RP SEQUENCE FROM N.A.
RA Gu Y.Z., Moran S.M., Hogenesch J.B., Wartman L., Bradfield C.A.;
RT "Molecular characterization and chromosomal localization of a third
RT alpha-class hypoxia induced factor subunit, HIF3alpha.";
RL Gene Expr. 0.0-0(1998).
DR EMBL; AF060194; AAC72734.1; -.
DR MGD; MGI:1859778; Hif3a.
DR InterPro; IPR001092; HLH_basic.
DR InterPro; IPR001067; Nuc_translocat.
DR InterPro; IPR001610; PAC.
DR InterPro; IPR000014; PAS_domain.
DR Pfam; PF00989; PAS; 2.
DR PRINTS; PR00785; NCTRNSLOCATR.
DR SMART; SM00353; HLH; 1.
DR SMART; SM00086; PAC; 1.
DR SMART; SM00091; PAS; 2.
DR PROSITE; PS00038; HELIX_LOOP_HELIX; UNKNOWN_1.
SQ SEQUENCE 662 AA; 73013 MW; 58740A1B6993D3B5 CRC64;
```

```
Alignment Scores: 7.93e-84 Length: 662
Pred. No.: 1045.00 Matches: 218
Score: 85.88% Conservative: 1
Percent Similarity: 85.49% Mismatches: 5
Best Local Similarity: 51.99% Indels: 31
Query Match: 11 Gaps: 3
DB:
US-09-896-791B-2 (1-1100) x Q92215 (1-662)
QY 43 AGTGTCAACACGAGCTCGGGAAGAGAGTCCGGGAGCGCGCGGAGCGGCGGAGC 102
Db 7 ArgSerAsnThrGluLeuArgLysGluLysSerArgAspAlaAlaArgSerArgSer 26
QY 103 CAGGAGAGCGAGGTCTGTACCAGCTGGCGCACACTGCGCCCTTTGCGCGCGGCTCAGC 162
Db 27 GlnGluThrGluValLeuTyrGlnLeuAlaHisThrLeuProPheAlaArgGlyValSer 46
QY 163 GCGCACCTGGACAAGCCCTCCATCATGCGCCTCACAAATCAGCTACCTGCGCATGCCGC 222
Db 47 AlaHisLeuAspLysAlaSerIleMetArgLeuThrIleSerTyrLeuArgMetHisArg 66
QY 223 CTCCTCCGACGAGTGGAA-----AAAGGGGAGAGCCACTGGAGCCCTG 267
Db 67 LeuCysAlaAlaGly-GluTrpAsnGlnValGluLysGlyGluProLeuAspAlaCy 86
QY 268 CTACCTGAAGGCCCTGGAGGTTTCGTATGTTACTACCGCCGAGGAGACATGGCTTA 327
Db 86 sTyrLeuLysAlaLeuGluGlyPheValMetValLeuThrAlaGluGlyAspMetAlaTy 106
QY 328 CCTGTGCGAAATGTCAGACACCTGGGCTCAGTCAGTGACCTCTGTTCCCTCC 387
Db 106 rLeuSerGluAsnValSerLysHisLeuGlyLeuSerGln----- 119
QY 388 CTGATACATAACCCCACTCTGGTACCAATTTCTCTGAGACTCATGGACACAGTATC 447
Db 120 -----Leu-GluLeuIleGlyHisSerIle 127
QY 448 TTTGATTTATCCATCCCTGTGACCAAGAGGAACTTCAAGAGCCCTGACCCCGAGCCG 507
Db 128 PheAspPheIleHisProCysAspGlnGluLeuGlnAspAlaLeuThrProArgPro 147
QY 508 AACCTGTCAAAGAAGAAGCTGGAAGCCCAACAGAGCGCCACTTTTCCCTGCGGAATGAAG 567
Db 148 AsnLeuSerLysLysLeuGluAlaProThrGluArgHisPheSerLeuArgMetLys 167
QY 568 AGCAGCTCACACGAGAGGGCCGACGCTCAACCTCAAGCGGCCACTTCCCTGCGAGGTGCTG 627
Db 168 SerThrLeuThrSerArgGlyArgThrLeuAsnLeuLysAlaAlaThrTrpLysValLeu 187
QY 628 CACTGCTCAGGACATATGAGGGCCCTACAGCCCTGCAATGCTTATCTGTGAAGCCATCCC----- 741
Db 208 ProArgSerGluProProLeuGlnCysLeuValLeuIleCysGluAlaIleProHisPro 227
QY 742 -----CAGCTCCCTTCCACGATGGTGCTACTCTG 771
Db 228 AlaSerLeuGluProProLeuGlyArgGlyAlaPheLeu 240
RESULT 4
Q8WXAL PRELIMINARY; PRT; 632 AA.
ID Q8WXAL
AC Q8WXAL;
DT 01-MAR-2002 (TReMBLrel. 20, Created)
DT 01-MAR-2002 (TReMBLrel. 20, Last sequence update)
DE 01-JUN-2002 (TReMBLrel. 21, Last annotation update)
DE Inhibitory PAS domain protein.
GN IPAS.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
```

OC Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Cheng J.Q.;  
RT "Cloning and characterization of human inhibitory PAS domain protein."  
RL Submitted (DEC-2001) to the EMBL/GenBank/DDBJ databases.  
DR EMBL; AF463492; AAL69947.1; -;  
DR InterPro; IPR001092; HLH\_basic.  
DR Pfam; PF00989; PAS; 2.  
DR SMART; SM00353; HLH; 1.  
DR SMART; SM00091; PAS; 2.  
DR PROSITE; PS00038; HELIX\_LOOP\_HELIX; UNKNOWN\_1.  
DR PROSITE; PS0112; PAS; 2.  
SQ SEQUENCE 632 AA; 68964 MW; 9665B0AF3998F8EF CRC64;

Alignment Scores:  
Pred. No.: 3.2e-81 Length: 632  
Score: 1015.50 Matches: 212  
Percent Similarity: 85.16% Conservative: 6  
Best Local Similarity: 82.81% Mismatches: 8  
Query Match: 50.52% Indels: 30  
DB: 4 Gaps: 3

US-09-896-791B-2 (1-1100) x Q8WXA1 (1-632)

QY 19 ATGGCGTTGGGCTGCAGCGCGTGGAGTGCACACCGAGCTGGGGAAGAGAGTCCGG 78  
|||||  
Db 1 MetAlaLeuGlyLeuGlnArgAlaArgSerThrThrGluLeuArgLysGluLysSerArg 20  
QY 79 GAGCGCGCGCGAGCGCGAGCCAGGAGAGGAGTCTGTACCAGCTGGCGCACACT 138  
|||||  
Db 21 AspaAlaArgSerArgSerGlnGluThrGluValLeuGluGlyPheValMetValLe 40  
QY 139 CTGCCCTTTTGGCGCGCTACCTGCAGCGCACCTGGACAGCGCTCCATCATGGCCCTCACA 198  
|||||  
Db 41 LeuProPheAlaArgGlyValSerAlaHisLeuAspLysAlaSerIleMetArgLeuThr 60  
QY 199 ATCAGCTACTCGGATGCACCGCTCTCGCAGCAGGT-----GGAAAA 243  
|||||  
Db 61 IleSerTyLeuArgMetHisArgLeuGlnValGlyGluTrpAsnGlnValGly-Al 80  
QY 244 AGGGGAGAGCCACTGCAGCGCTGTACCTGAGAGCGCTGGAGGTTTGTCTATGCTACT 303  
|||||  
Db 80 aGlyGlyGluProLeuAspAlaCysTyLeuLysAlaLeuGluGlyPheValMetValLe 100  
QY 304 CACCGCGAGGAGACATGGCTTACCTGTGGAAAAATGTGACAGACACTGGCGCTCAG 363  
|||||  
Db 100 uThrAlaGluGlyAspMetAlaTyLeuSerGluAsnValSerLysHisLeuGlyLeu 120  
QY 364 TCAGTGGAGCTCTGTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 423  
|||||  
Db 120 rGln-----Le 122  
QY 424 CTGGAGCTCATGGACACAGTATCTTTGATTTTATCCATCCCTGTGACCAAGAGGAACCTT 483  
|||||  
Db 122 u-GluLeuIleGlyHisSerIlePheAspPheIleHisProcysAspGlnGluGluLeu 141  
QY 484 CRAAGCGCCTGACCCCGCGCGAGACCTGTCAAGAAAGAGTGGAAAGCCCAACAGAG 543  
|||||  
Db 142 GlnAspAlaLeuThrProGlnGlnThrLeuSerArgLysValGluAlaProThrGlu 161  
QY 544 CGGCACCTTTTCCGTGGATGAGAGCAGCTCACAGAGAGGCGCAGCTCAACCTC 603  
|||||  
Db 162 ArgCysPheSerLeuArgMetLysSerThrLeuThrSerArgGlyArgThrLeuAsnLeu 181  
QY 604 AAAGCGGCGCACCTGGAGGTGTGCACTGCTCAGGACATATGAGGCGCTACAGGCCCTCT 663  
|||||  
Db 182 LysAlaIleThrTrpLysValLeuAsnCysSerGlyHisMetArgAlaTyLeuLysProPro 201  
QY 664 GCACAGACTTCCCTGCCGGGAGCCCTCGCTCCGAGCCCTCCCTGCAATGCCTGTGCTT 723

Db 202 AlaGlnThrSerProAlaGlySerProAspSerGluProGluProLeuGlnCysLeuValLeu 221  
|||||  
QY 724 ATCTGTGAAGCCATCCCGCAGCTCCCTTCCAGTAGGTGCT 765  
|||||  
Db 222 IleCysGluAlaIlePro-----HisProGlySer 231  
RESULT 5  
Q96K34  
ID Q96K34 PRELIMINARY; PRT; 632 AA.  
AC Q96K34;  
DT 01-DEC-2001 (TREMBlrel. 19, Created)  
DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)  
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)  
DE CDNA FLJ14819 fis, clone OVARC100241, moderately similar to  
DE hypoxia-inducible factor 1 alpha.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=OVARIAN CARCINOMA;  
RA Isogai T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,  
RA Nishikawa T., Nagai K., Sugano S., Ishibashi T., Fujimori K.,  
RA Tanai H., Kimata M., Watanabe M., Hiraoka S., Ishii S., Kawai Y.,  
RA Saito K., Yamamoto J., Wakamatsu A., Nakamura Y., Nagahari K.,  
RA Masuho Y., Kanehori K.;  
RT "NEDO human cDNA sequencing project."  
RL Submitted (MAY-2001) to the EMBL/GenBank/DDBJ databases.  
DR EMBL; AK027725; BAB55324.1; -;  
DR InterPro; IPR001092; HLH\_basic.  
DR InterPro; IPR001610; PAC.  
DR InterPro; IPR000014; PAS\_domain.  
DR Pfam; PF00989; PAS; 2.  
DR SMART; SM00086; PAC; 1.  
DR PROSITE; PS00038; HELIX\_LOOP\_HELIX; UNKNOWN\_1.  
SQ SEQUENCE 632 AA; 68934 MW; A19F1ED3D05E7A71 CRC64;

Alignment Scores:  
Pred. No.: 7.23e-81 Length: 632  
Score: 1011.50 Matches: 211  
Percent Similarity: 84.77% Conservative: 6  
Best Local Similarity: 82.42% Mismatches: 9  
Query Match: 50.32% Indels: 30  
DB: 4 Gaps: 3

US-09-896-791B-2 (1-1100) x Q96K34 (1-632)

QY 19 ATGGCGTTGGGCTGCAGCGCGTGGAGTGCACACCGAGCTGGGGAAGAGTCCGG 78  
|||||  
Db 1 MetAlaLeuGlyLeuGlnArgAlaArgSerThrThrGluLeuArgLysGluLysSerArg 20  
QY 79 GAGCGCGCGCGAGCGCGAGCCAGGAGCGAGGTGCTGTACCAGTGGCGCACACT 138  
|||||  
Db 21 AspaAlaArgSerArgSerGlnGluThrGluValLeuGluGlyPheValMetValLe 40  
QY 139 CTGCCCTTTTGGCGCGCTACCTGCAGCGCACCTGGACAGCGCTCCATCATGGCCCTCACA 198  
|||||  
Db 41 LeuProPheAlaArgGlyValSerAlaHisLeuAspLysAlaSerIleMetArgLeuThr 60  
QY 199 ATCAGCTACTCGGATGCACCGCTCTCGCAGCAGGT-----GGAAAA 243  
|||||  
Db 61 IleSerTyLeuArgMetHisArgLeuGlnValGlyGluTrpAsnGlnValGly-Al 80  
QY 244 AGGGGAGAGCCACTGGAGCGCTGTACCTGAGAGCGCTGGAGGTTTGTCTATGCTACT 303  
|||||  
Db 80 aGlyGlyGluProLeuAspAlaCysTyLeuLysAlaLeuGluGlyPheValMetValLe 100  
QY 304 CACCGCGAGGAGACATGGCTTACCTGTGGAAAAATGTGACAGACACTGGCGCTCAG 363  
|||||  
Db 100 uThrAlaGluGlyAspMetAlaTyLeuSerGluAsnValSerLysHisLeuGlyLeu 120





OC Mammalia; Euthera; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
OX NCBI\_TaxID=10116;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=LIVER;  
RX MEDLINE=21134367; PubMed=11237857;  
RA Kietzmann T., Cornesse Y., Brechtel K., Modaresi S., Jungermann K.;  
RT "Perivenous expression of the mRNA of the three hypoxia-inducible  
RT factor a-subunits HIF-1a, HIF2a and HIF3a in rat liver.";  
RL Biochem. J. 354:531-537(2001).  
DR EMBL: AJ277827; CAB96611.1; .  
DR InterPro: IPR001092; HLH\_basic.  
DR InterPro: IPR001067; Nuc\_translocat.  
DR InterPro: IPR001610; PAC.  
DR Pfam: PF00989; PAS; 2.  
DR PRINTS: PR00785; NCTRNLOCATR.  
DR SMART: SM00353; HLH; 1.  
DR SMART: SM00086; PAC; 1.  
DR SMART: SM00091; PAS; 2.  
DR PROSITE: PS00038; HELIX\_LOOP\_HELIX; UNKNOWN\_1.  
SQ SEQUENCE 662 AA; 72887 MW; AC9672E340544010 CRC64;

Alignment Scores:  
Pred. No.: 6.19e-80 Length: 662  
Score: 1001.00 Matches: 211  
Percent Similarity: 83.92% Conservative: 3  
Best Local Similarity: 82.75% Mismatches: 10  
Query Match: 49.80% Indels: 31  
DB: 11 Gaps: 3

US-09-896-791B-2 (1-1100) x Q9JHS2 (1-662)

QY 43 AGCTCGAACACCGAGTCCGGAGGAGAGTCCGGAGCGCGCCGCGAGCGCGCGCAGC 102  
Db 7 ArgSerThrGluLeuArgLysGluLysSerArgAspAlaAlaArgSerArgSer 26  
QY 103 CAGGACGAGGTGCTGACACGCTGGCGCACACTCTGCTTTCGCGCGCGCGTCAGC 162  
Db 27 GlnGluThrGluValLeuTyrGlnLeuAlaHisThrLeuProPheAlaArgGlyValSer 46  
QY 163 GCGCACCTGGACAGCGCTCCATCATGCGCTCACAAATCAGTACCTGCGCATGACCGC 222  
Db 47 AlaHisLeuAspLysAlaSerIleMetArgLeuThrIleSerTyrLeuArgMetHisArg 66  
QY 223 CTCTGCGCAGCAGTGGAA-----AAAGGGGAGAGCCACTGGACGCCCTG 267  
Db 67 LeuCyAlaAlaGly-GluTrpAsnGlnValArgLysGluGlyGluProLeuAspAlaCy 86  
QY 268 CTACCTGAAGCGCCTGGAGGGTTTCGTATGTTACTACCGCGCGAGGAGACATGGCTTA 327  
Db 86 sTyrLeuLysAlaLeuGluGlyPheValMetValLeuThrAlaGluGlyAspMetAlaTy 106  
QY 328 CTGTGCGGAATATGTCAGAACGACCTGGCGCTCAGTACGTAGTGACCTCTGTTCTCTCC 387  
Db 106 rLeuSerGluAsnValSerLysHisLeuGlyLeuSerGln----- 119  
QY 388 CTGATACATAACCCCACTCTCTGTGTACCAATTTCTCTGAGTCAATTCGACACAGTATC 447  
Db 120 -----Leu--GluLeuIleGlyHisSerIle 127  
QY 448 TTTGATTTTATCCATCCCTGTGACCAAGAGACTTCAAGAGCGCTGACCCCGAGCGCG 507  
Db 128 PheAspPheIleHisProCysAspGlnGluLeuGlnAspAlaLeuThrProArgPro 147  
QY 508 AACCTGTCAAAGAGAGTGGAGGCCCAACAGAGCGCCACTTTTCCTCGCAATGAG 567  
Db 148 SerLeuSerLysLysSerGluAlaAlaThrGlyArgHisPheSerLeuArgMetLys 167  
QY 568 AGCACGCTCAGCAGAGCGCGCAGCTCAACCTCAAGCGGCCACCTGGAAGGTGCTG 627  
Db 168 SerThrLeuThrSerArgGlyArgAlaLeuAsnLeuLysAlaAlaThrTrpLysValLeu 187

QY 628 CACTGCTCAGACATATGAGGCGCTACAGCCCGCTGCACAGACTTCCCTCGCGGAGC 687  
Db 188 HisCysSerGlyHisMetArgAlaTyrLysProProAlaGlnThrSerProAlaGlySer 207  
QY 688 CQTCGCTCGAGCCTCCCTGCAATGCTGCTGCTTATCTGTGAAGCCATCCCC----- 741  
Db 208 ProArgSerGluProLeuGlnCysLeuValLeuIleCysGluAlaIleProHisPro 227  
QY 742 -----CAGTCCCTCCCTCCACGATGCTGTACTCTG 771  
Db 228 AlaSerLeuGluProProLeuGlyArgGlyAlaPheLeu 240

RESULT 8  
Q9Y2N7  
ID Q9Y2N7 PRELIMINARY; PRT; 667 AA.  
AC Q9Y2N7;  
DT 01-NOV-1999 (TREMBLrel. 12, Created)  
DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)  
DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)  
DE Putative homolog of hypoxia inducible factor three alpha  
DE (Hypoxia-inducible factor-3 alpha).  
GN HIF-3A.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Lamerdin J.E., McCreedy P.M., Skowronski E., Viswanathan V.,  
RA Burkhardt-Schultz K.J., Gordon L., Dias J., Ramirez M., Stillwagen S.,  
RA Phan H., Velasco N., Do L., Regala W., Terry A., Garnes J.,  
RA Danganan L., Erler A., Christensen M., Georgescu A., Avila J., Liu S.,  
RA Attix C., Andreise T., Trankheim M., Amico-Keller G., Coefield J.,  
RA Duarte S., Lucas S., Bruce R., Thomas P., Quan G., Kronmiller B.,  
RA Arellano A., Sanders C., Ow D., Nolan M., Trong S., Kobayashi A.,  
RA Olsen A.S., Carrano A.V.;  
RT "Sequence analysis of a 1.9 Mb region in 19q13.2 between APOE and  
RT D19S412.";  
RL Submitted (APR-1999) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP SEQUENCE FROM N.A.  
RC TISSUE=KIDNEY;  
RA Hara S., Kobayashi C., Hamada J., Imura N.;  
RT "Characterization of human hypoxia-inducible factor-3 alpha.";  
RL Submitted (JAN-2001) to the EMBL/GenBank/DBJ databases.  
DR EMBL: AC007193; RAD22668.1; .  
DR EMBL: AB054067; BAB69689.1; .  
DR InterPro: IPR001092; HLH\_basic.  
DR InterPro: IPR001067; Nuc\_translocat.  
DR InterPro: IPR001610; PAC.  
DR Pfam: PF00989; PAS; 2.  
DR PRINTS: PR00785; NCTRNLOCATR.  
DR SMART: SM00353; HLH; 1.  
DR SMART: SM00086; PAC; 1.  
DR SMART: SM00091; PAS; 2.  
DR PROSITE: PS00038; HELIX\_LOOP\_HELIX; UNKNOWN\_1.  
SQ SEQUENCE 667 AA; 72404 MW; 67B8794FF9DCCF4B CRC64;

Alignment Scores:  
Pred. No.: 2.68e-78 Length: 667  
Score: 982.50 Matches: 205  
Percent Similarity: 85.08% Conservative: 6  
Best Local Similarity: 82.65% Mismatches: 7  
Query Match: 48.88% Indels: 30  
DB: 4 Gaps: 3

US-09-896-791B-2 (1-1100) x Q9Y2N7 (1-667)

QY 43 AGCTGAACACCGAGTCCGGAGGAGAGTCCGGAGCGCGCCGCGAGCGCGCAGC 102  
Db 7 ArgSerThrThrGluLeuArgLysGluLysSerArgAspAlaAlaArgSerArgSer 26



DT	01-JUN-2002 (TReMBL)rel. 21, Last annotation update)
DE	Hypoxia inducible factor 1, alpha subunit.
GN	HIF1A.
OS	Mus musculus (Mouse).
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
ON	NCBI_TaxID=10090;
FX	[1]
RP	SEQUENCE FROM N.A.
RC	STRAIN=C57BL/6J; TISSUE=EMBRYO;
RC	MEDLINE=21085660; PubMed=11217851;
RA	Kawai J., Shingawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA	Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA	Aizawa K., Izawa M., Nishi K., Kiyosawa H., Konno S., Yamanaka I.,
RA	Saito T., Okazaki Y., Gojibori T., Bono H., Kasukawa T., Saito R.,
RA	Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA	Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA	Kuehl P., Lewis S., Matsuo Y., Nikaide I., Pesole G., Quackenbush J.,
RA	Schraml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA	Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA	Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA	Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA	Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA	Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA	Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA	Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA	Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA	Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,
RA	Hayashizaki Y.;
RT	"Functional annotation of a full-length mouse cDNA collection.";
RL	Nature 409:685-690(2001).
DR	EMBL; AK017853; BAB30975.1; -.
DR	MGD; MGI:106918; Hifla.
DR	InterPro: IPR001092; HLH_basic.
DR	InterPro: IPR001067; Nuc_translocat.
DR	InterPro: IPR000014; PAS_domain.
DR	Pfam; PF00989; PAS; 1.
DR	PRINTS; PR00785; NCTRNLOCATR.
DR	SMART; SM00353; HLH; 1.
DR	SMART; SM00091; PAS; 1.
DR	PROSITE; PS00038; HELIX_LOOP_HELIX; UNKNOWN_1.
SQ	SEQUENCE 258 AA; 29703 MW; C42D1CD68117C552 CRC64;
Alignment scores:	
Pred. No.:	9,82e-46
Score:	613.50
Length:	258
Matches:	129
Percent Similarity:	67.49%
Conservative:	35
Best Local Similarity:	53.09%
Mismatches:	51
Query Match:	Indels: 28
DB:	Gaps: 3
US-09-896-791B-2 (1-1100) x Q9CYA8 (1-258)	
Qy	34 CAGCGCGTAGGTCGACACCGAGCTCGGAGGAGAGTCCGGGAGCGCGCCGCGAGC 93
Db	9 GluLysLysLysMetSerGluArgLysGluLysSerArgAspAlaAlaArgSer 28
Qy	94 CGCGGAGCCAGCAGGAGGAGGTCGTACACGCTGGCGCACACTCGCCCTTTGCGCGC 153
Db	29 ArgArgSerLysGluSerGluValPheTyrGluLeuAlaHisGlnLeuProLeuProHis 48
Qy	154 GGGCTCAGCGCGCACCTGGACAGGCGCTCCATCATCGCGCTCACCAATCAGCTACGCTGCGC 213
Db	49 AsnValSerSerHisLeuAspLysAlaSerValMetArgLeuThrIleSerTyrLeuArg 68
Qy	214 ATGCACCGCCTCTCGCAGCAGCGAGTGGAAAGGGGGAGAG-----CCACT 258
Db	69 ValArgLysLeuLeuAspAlaGlyGly-LeuAspSerGluAspGluMetLysAlaGlnMe 88
Qy	259 GGAGCGCTGCTACTCAAGCGCCTCGAGGGTTTCGTATGCTACTGCTACCGCGGAGGAGA 318
Db	88 tAspCysPheTyrLeuLysAlaLeuAspGlyPheValMetValLeuThrAspAspGlyAs 108

```
QY 319 CATGCTTACCTGTCGGAATAATGTGACGAAGCACTGGCGCTCAGTCAGTGACCTCTGT 378
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 108 pMetValTyrIleSerAspAsnValAsnLysTyrMetGlyLeuThrGln----- 124

QY 379 TCCTCTCCCTGATACATACCCCACTCCTGGTACCAATTTCTCTCTGGAGCTCATTTGA 438
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 125 -----Phe--GluLeuThrGly 129

QY 439 CACATATCTTTGATTTCATCCCTGTGACCAAGAGAACTTCAAGACGCCCTGACC 498
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 130 HisSerValPheAspPheThrHisProCysAspHisGluMetArgGluMetLeuThr 149

QY 499 CCCAGGCCCACTCTCAAGAAGAAGCTGGAAGCCCAACAGACGCCACTTTTCCCTG 558
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 150 HisArgAsnGlyProValArgLysGlyLysGluLeuAsnThrGlnArgSerPheLeu 169

QY 559 CGAATGAGAGACACCTCACAGACAGAGCGCGCAGCTCAACCTCAAGCGGCCACCTGG 618
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 170 ArgMetLysCysThrLeuThrSerArgGlyArgThrMetAsnIleLysSerAlaThrTrp 189

QY 619 AAGGTGCTGCACTGCTCAGGACATATGAGGGCTCAAGCCCTCCAGACACTTCCCT 678
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 190 LysValLeuHisCysThrGlyHisIleHisValTyrAspThrAsnSerAsnGlnProGln 209

QY 679 GCCGGGAGCCCTCGCTCCGAGCCTCCCTGCAATGCCTGGTGTATCTGTGAAGCCATC 738
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 210 CysGly-----TyrLysLysProPromethrCysLeuValLeuIleCysGluProIle 227

QY 739 CCC 741
    |||||
Db 228 Pro 228

RESULT 12
Q8R385
ID Q8R385 PRELIMINARY; PRT; 836 AA.
AC Q8R385;
DT 01-JUN-2002 (TremBLrel. 21, Created)
DT 01-JUN-2002 (TremBLrel. 21, Last sequence update)
DT 01-JUN-2002 (TremBLrel. 21, Last annotation update)
DE Hypoxia inducible factor 1, alpha subunit.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Strausberg R.;
RL Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC026139; AAH26139.1; -.
SQ SEQUENCE 836 AA; 93469 MW; 8A7B21B0F13E84D5 CRC64;
```

```
Alignment Scores:
Pred. No.: 1.23e-45 Length: 836
Score: 613.50 Matches: 129
Percent Similarity: 67.49% Conservative: 35
Best Local Similarity: 53.09% Mismatches: 51
Query Match: 30.52% Indels: 28
DB: 11 Gaps: 3

US-09-896-791B-2 (1-1100) x Q8R385 (1-836)

QY 34 CAGCGCGTGAGTCCGACACCGAGCTGCGGAAGGAGAGTCCGGGAGCGCGCCGCGAGC 93
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 9 GluLysLysLysMetSerSerGluArgLysGluLysSerArgAspAlaAlaArgSer 28

QY 94 CGCGCAGCCAGGAGAGGTGTGTACCAAGTGGCGCACACTCTGCCCTTTGCGGCGC 153
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 29 ArgArgSerLysGluSerGluValPheTyrGluLeuAlaHisGlnLeuProLeuProHis 48

QY 154 GCGTCACGCGGCACCTGGACAGGCCCTTCATCATCGCGCTCAACAATCAGCTACCTGGCG 213
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 49 AsnValSerSerHisLeuAspLysAlaSerValMetArgLeuThrIleSerTyrLeuArg 68
```

```
QY 214 ATGCACCGCCTCTGCGCAGCAGCTGGAAAAGGGGAGAG-----CCACT 258
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 69 ValArgLysLeuLeuAspAlaGlyGly-LeuAspSerGluAspGluMetLysAlaGlnMe 88

QY 259 GGACGCCCTGCTACCTGAAGGCCCTGGAGGTTTGTGTCATGGTACTACACCGCGAGGAGA 318
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 88 TaspCysPheTyrLeuLysAlaLeuAspGlyPheValMetValLeuThrAspAspGlyAs 108

QY 319 CATGCTTACCTGTCGGAATAATGTGACGAAGCACTGGCGCTCAGTCAGTGCACCTCTGT 378
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 108 pMetValTyrIleSerAspAsnValAsnLysTyrMetGlyLeuThrGln----- 124

QY 379 TCCTCTCCCTGATACATACCCCACTCCTGGTACCAATTTCTCTCTGGAGCTCATTTGA 438
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 125 -----Phe--GluLeuThrGly 129

QY 439 CACATATCTTTGATTTCATCCCTGTGACCAAGAGAACTTCAAGACGCCCTGACC 498
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 130 HisSerValPheAspPheThrHisProCysAspHisGluMetArgGluMetLeuThr 149

QY 499 CCCAGGCCCACTCTCAAGAAGAAGCTGGAAGCCCAACAGACGCCACTTTTCCCTG 558
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 150 HisArgAsnGlyProValArgLysGlyLysGluLeuAsnThrGlnArgSerPheLeu 169

QY 559 CGAATGAGAGACACCTCACAGACAGAGCGCGCAGCTCAACCTCAAGCGGCCACCTGG 618
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 170 ArgMetLysCysThrLeuThrSerArgGlyArgThrMetAsnIleLysSerAlaThrTrp 189

QY 619 AAGGTGCTGCACTGCTCAGGACATATGAGGGCTCAAGCCCTCCAGACACTTCCCT 678
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 190 LysValLeuHisCysThrGlyHisIleHisValTyrAspThrAsnSerAsnGlnProGln 209

QY 679 GCCGGGAGCCCTCGCTCCGAGCCTCCCTGCAATGCCTGGTGTATCTGTGAAGCCATC 738
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 210 CysGly-----TyrLysLysProPromethrCysLeuValLeuIleCysGluProIle 227

QY 739 CCC 741
    |||||
Db 228 Pro 228

RESULT 13
Q9XTA5
ID Q9XTA5 PRELIMINARY; PRT; 823 AA.
AC Q9XTA5;
DT 01-NOV-1999 (TremBLrel. 12, Created)
DT 01-NOV-1999 (TremBLrel. 12, Last sequence update)
DT 01-MAR-2002 (TremBLrel. 20, Last annotation update)
DE Hypoxia-inducible factor-1 alpha.
GN HIF-1 ALPHA.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Bovinae; Bos.
OX NCBI_TaxID=9913;
RN [1]
RP SEQUENCE FROM N.A.
RA Hara S., Kobayashi C., Imura N.;
RT "Molecular cloning of cDNAs encoding hypoxia-inducible factor (HIF)-
RL Biochim. Biophys. Acta 1445:237-243(1999).
DR EMBL; AB018398; BAA78675.1; -.
DR InterPro; IPR001092; HLH_basic.
DR InterPro; IPR001321; HypoxindfIA.
DR InterPro; IPR001610; PAC.
DR InterPro; IPR000014; PAS_domain.
DR Pfam; PF00785; PAC; 1.
DR Pfam; PF00989; PAS; 2.
DR PRINTS; PR01080; HYPOXIAFIA.
DR SMART; SM00353; HLH; 1.
DR SMART; SM00086; PAC; 1.
DR SMART; SM00091; PAS; 2.
```

DR PROSITE; PS00038; HELIX\_LOOP\_HELIX; UNKNOWN\_1.  
SQ SEQUENCE 823 AA; 92127 MW; 12674E467A61B1A1 CRC64;

## Alignment Scores:

Pred. No.: 1-84e-45 Length: 823  
Score: 611.50 Matches: 130  
Percent Similarity: 68.49% Conservative: 33  
Best Local Similarity: 54.62% Mismatches: 47  
Query Match: 30.42% Indels: 28  
DB: 3 Gaps: 3

US-09-896-791B-2 (1-1100) x Q9YIB9 (1-823)

QY 49 AACACCGAGCTCGGAGGAGAGAGTCCGGGACGCGCGCGCGCGCGCGCGCGCGCGAGG 108  
Db 14 SerSerGluArgArgLysGluLysSerArgAspAlaAlaArgSerArgArgSerLysGlu 33  
QY 109 ACGAGGTGCTACACAGTGGCGCACACTCTGCGCTTTCGCGCGCGCGGTGACGCGCAC 168  
Db 34 SerGluValPheTyrGluLeuAlaHisGlnLeuProLeuProHisAsnValSerSerHis 53  
QY 169 CTGCAAGCGCTCCATCATCGCCCTCACAAATCAGCTACCTGCGCATGCGACCGCTCTGC 228  
Db 54 LeuAspLysAlaSerValMetArgLeuThrIleSerTyrLeuArgValArgLysLeuLeu 73  
QY 229 GCAGCAGGTGGAAAGGGGAGAG-----CCACTGGACGCGCTGTACCT 273  
Db 74 AspAlaGly-AspLeuAspIleGluAspGluMetLysAlaGlnMetAsnCysPheTyrLe 93  
QY 274 GAAGGCGCTGGAGGTTTCCTCATGTGACTACCGCGGAGGAGACATGCTTACCTGTG 333  
Db 93 uLysAlaLeuAspGlyPheValMetValLeuThrAspAspGlyAspMetIleTyrIleSe 113  
QY 334 GAAATATGTCAGCAAGCACTGCGCTCAGTCAGTGGACCTGTCTCTCTCTCTCTCTCT 393  
Db 113 rAspAsnValAsnLysTyrMetGlyLeuThrGln----- 124  
QY 394 CATACCCCACTCCTGGTACCAATTTCTCTCTGGAGCTCATTTGGACACATATCTTTGAT 453  
Db 125 -----Phe--GluLeuThrGlyHisSerValPheAsp 134  
QY 454 TTTATCATCTCTGTGACCAAGAGGAACTTCAAGACGCGCTGACCGCGGAGCGCAACCTG 513  
Db 135 PheThrHisProCysAspHisGluMetArgLysGluMetArgLysHisArgGlyLeu 154  
QY 514 TCAAGAAGAAGCTGGAAGCCCAACAGACGCGCACTTTTCCCTGCGCAATGAAGACGACG 573  
Db 155 valLysLysGlyLysGluGlnAsnThrGlnArgSerPhePheLeuArgMetLysCysThr 174  
QY 574 CTCACGACGAGGCGGCGCGCTCAACCTCAAGCGCGCCACCTGGAAGGTGCTGCGACTGC 633  
Db 175 LeuThrSerArgGlyArgThrMetAsnIleLysSerAlaThrTrpLysValLeuHisCys 194  
QY 634 TCAGGACATATAGGCGCTACAGCCCGCTGACAGACTTCCCTCGCGGAGCCCTCGC 693  
Db 195 ThrGlyHisIleHisValTyrAspThrAsnSerAsnGlnSerGlnCysGly-----Tyr 212  
QY 694 TCGAGCTCTCCCTGCAATGCTGTGCTTATCTGTGAGCCATCCCC 741  
Db 213 LysLysProPromethrCysLeuValLeuIleCysGluProIlePro 228

## RESULT 14

Q9YIB9  
ID Q9YIB9 PRELIMINARY; PRT; 811 AA.  
AC Q9YIB9;  
DT 01-MAY-1999 (TrEMBLrel. 10, Created)  
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)  
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)  
DE Hypoxia-inducible factor-1 alpha.  
OS Gallus gallus (Chicken).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;  
OC Gallus.

OX NCBI\_TaxID=9031;  
RN [1]  
RP SEQUENCE FROM N.A.

RC TISSUE=HEART;  
RA Takahashi T.;

RT "Molecular cloning and expression of an avian cDNA for hypoxia-  
inducible factor-1 alpha in embryonic ventricular myocytes.";  
RL Submitted (MAY-1998) to the EMBL/GenBank/DBJ databases.

DR EMBL; AB013746; BAA34234.2; -;  
DR InterPro; IPR001092; HLH\_basic.

DR InterPro; IPR001610; PAC.

DR InterPro; IPR000014; PAS\_domain.

DR Pfam; PF00785; PAC; 1.

DR Pfam; PF00989; PAS; 2.

DR SMART; SM00353; HLH; 1.

DR SMART; SM00086; PAC; 1.

DR SMART; SM00091; PAS; 2.

DR PROSITE; PS00038; HELIX\_LOOP\_HELIX; UNKNOWN\_1.

SQ SEQUENCE 811 AA; 90542 MW; D14CD9FC9BF064CB CRC64;

## Alignment Scores:

Pred. No.: 4-14e-45 Length: 811  
Score: 607.50 Matches: 136  
Percent Similarity: 67.36% Conservative: 25  
Best Local Similarity: 56.90% Mismatches: 51  
Query Match: 30.22% Indels: 28  
DB: 13 Gaps: 3

US-09-896-791B-2 (1-1100) x Q9YIB9 (1-811)

QY 43 AGTTCGAACACCGAGCTCGGAGGAGAGTCCGGGACGCGCGCGCGCGCGCGCGCAGC 102  
Db 12 ArgIleSerSerGluArgArgLysGluLysSerArgAspAlaAlaArgCysArgSer 31  
QY 103 CAGGACACGAGGTGCTGTACACAGTGGCGCACACTCTGCGCTTTCGCGCGCGCTCAGC 162  
Db 32 LysGluSerGluValPheTyrGluLeuAlaHisGlnLeuProLeuProHisThrValSer 51  
QY 163 GCCACCTGGACAAAGCGCTCATCATGCGCTCACAAATCAGCTACCTGCGCATGCGCGC 222  
Db 52 AlaHisLeuAspLysAlaSerIleMetArgLeuThrIleSerTyrLeuArgMetArgLys 71  
QY 223 CTCTGCGCAGAGTGGAAAGGGGAGAG-----CCACTGACGCGCTG 267  
Db 72 LeuLeuAspAlaGly-GluLeuGluThrGluAlaAsnMetGluLysGluLeuAsnCysPh 91  
QY 268 CTACCTGAAGCGCTCGAGGGTTTCGTCTGCTACTGCTACCGCGGAGGAGACATGCTTGA 327  
Db 91 eTyrLeuLysAlaLeuAspGlyPheValMetValLeuSerGluAspGlyAspMetIleTy 111  
QY 328 CCGTTCGGAATATGTGAGCAAGCACCTGGGCGCTCAGTCAGTG-GACCTCTGTCTCCTC 386  
Db 111 rMetSerGluAsnValAsnLysCysMetGlyLeuThrGlnPheAsp----- 126  
QY 387 CCTGATACATACCCCACTCTCTGGTACCAATTTCTCTCTGAGCTCATTTGGACACAGTAT 446  
Db 127 -----LeuThrGlyHisSerVa 132  
QY 447 CTTTGATTTTATCCATCCCTGTGTCACCAAGAGGAACTTCAAGACGCGCTGACCCCGAGGCC 506  
Db 132 lPheAspPheThrHisProCysAspHisGluGluLeuArgGluMetLeuThrHisArgAs 152  
QY 507 GAACCTGTCAAGAAGAAGCTGGAAGCCCAACAGACGCGCACTTTTCCCTGCGCAATGAA 566  
Db 152 nGlyProValLysLysGlyLysGluGlnAsnThrGluArgSerPhePheLeuArgMetLy 172  
QY 567 GAGCAGCTCACACGAGGCGCGCGCTCAACCTCAAGCGCGCCACCTGGAAGTGTG 626  
Db 172 sCysThrLeuThrSerArgGlyArgThrValAsnIleLysSerAlaThrTrpLysValLe 192  
QY 627 GCACCTCTCAGGACATATGAGGCGCTTACAAAGCCCTGACAGACTTCCCTCGCGGAG 686  
Db 192 uHisCysThrGlyHisIleArgValTyrAspThrCysAsnAsnGlnThrHisCysGly-- 211

QY 687 CCCTGGCTCCGAGCCTCCCTGCAATGCTGCTGCTTATCTGTGAAGCCATCCCC 741  
Db 212 ---TyrLysLysProMetThrCysLeuValLeuIleCysGluProIlePro 228  
RESULT 15  
Q96PT9 ID Q96PT9 PRELIMINARY; PRT; 735 AA.  
AC Q96PT9; ID Q96PT9; PRT; 735 AA.  
DT 01-DEC-2001 (Tremblrel. 19, Created)  
DT 01-DEC-2001 (Tremblrel. 19, Last sequence update)  
DT 01-MAR-2002 (Tremblrel. 20, Last annotation update)  
DE Hypoxia-inducible factor 1 alpha variant.  
GN HIF1A.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=LIVER;  
RA Tanaka S., Sugimachi K.;  
RT "Hypoxia-inducible factor-1 alpha variant isolated from human liver tissue."  
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AB073325; BAB70608.1; -;  
DR InterPro; IPR001092; HLH\_basic.  
DR InterPro; IPR001610; PAC.  
DR InterPro; IPR000014; PAS\_domain.  
DR Pfam; PF00785; PAC; 1.  
DR Pfam; PF00989; PAS; 2.  
DR SMART; SM00086; PAC; 1.  
DR PROSITE; PS00038; HELIX\_LOOP\_HELIX; UNKNOWN\_1.  
SQ SEQUENCE 735 AA; 82746 MW; 34DD604FB4E4418E CRC64;

## Alignment Scores:

Pred. No.: 4.98e-45 Length: 735  
Score: 606.50 Matches: 129  
Percent Similarity: 68.07% Conservative: 33  
Best Local Similarity: 54.20% Mismatches: 48  
Query Match: 30.17% Indels: 28  
DB: 4 Gaps: 3

US-09-896-791B-2 (1-1100) x Q96PT9 (1-735)

QY 49 ACACCGAGCTCGGAGAGAGTCCGCGGAGCGCGCGCGCGCGCGCGCGCGAGGAG 108  
Db 14 SerSerGluArgArgLysLysSerArgAspAlaAlaArgSerArgSerLysGlu 33  
QY 109 ACGGAGGTCTCTACGAGTGGCGGACACTCTGCCCTTTTCGCGCGCGCGCGCGCAC 168  
Db 34 SerGluValPheTyrGluLeuAlaHisGlnLeuProLeuProHisAsnValSerHis 53  
QY 169 CTGGACAGCGCTCCATCATGCGCTCACAATCAGCTACTGCGCATGCGCGCGCGCTCTGC 228  
Db 54 LeuAspLysAlaSerValMetArgLeuThrIleSerTyrLeuArgValArgLysLeu 73  
QY 229 GCAGCAGTGGGAAAGGGGGAGAG-----CCACTGGAGCGCTGTCTACCT 273  
Db 74 AspAlaGly-AspLeuAspIleGluAspMetLysAlaGlnMetAsnValSerHis 93  
QY 274 GAAGCCCTGGAGGTTTCGTGATGCTACTACCGCGCGGAGGAGACATGGCTTACCTGTC 333  
Db 93 ulysAlaLeuAspGlyPheValMetValLeuThrAspAspGlyAspMetIleTyrIleSe 113  
QY 334 GGAATATGTCAGCAGCAGCTGGGCTCAGTCAGTGGACCTCTGTCTCTCCCTCCCTGATA 393  
Db 113 rAspAsnValAsnLysTyrMetGlyLeuThrGln----- 124  
QY 394 CATAACCCCATCTCTGGTACCAATTCTCTCTGGAGCTCATTTGGACACAGTATCTTGTAT 453  
Db 125 -----Phe--GluLeuThrGlyHisSerValPheAsp 134

Search completed: December 2, 2002, 20:39:17  
Job time : 70 secs

QY 454 TTTATCCATCCTGTGACCAAGAGGAACCTTCAAGACGCGCCTGACCCCGAGCGCAACCTG 513  
Db 135 PheThrHisProCysAspHisGluGluMetArgGluMetLeuThrHisArgAsnGlyLeu 154  
QY 514 TCAAAGAAGAAGCTGGAAGCCCCAACAGAGCGCGCCTTTTCCCTGCGAATGAAGACGACG 573  
Db 155 ValLysLysGlyLysGluGlnAsnThrGlnArgSerPhePheLeuArgMetLysCysThr 174  
QY 574 CTCACGACGAGCGGCGCACGCTCAACCTCAAAGCGCGCCACCTGGAGAGTGTGCACTGC 633  
Db 175 LeuThrSerArgGlyArgThrMetAsnIleLysSerAlaThrIrrLysValLeuHisCys 194  
QY 634 TCAGGACATATGAGGCGCTTACAAGCCCCCTGCACAGACTTCCCTCGCGGAGCCCTCGC 693  
Db 195 ThrGlyHisIleHisValTyrAspThrAsnSerAsnGlnProGlnCysGly-----Tyr 212  
QY 694 TCCGAGCCTCCCTCGAATGCTGGTGTATCTGTGAAGCCATCCCC 741  
Db 213 LysLysProProMetThrCysLeuValLeuIleCysGluProIlePro 228

**THIS PAGE BLANK (USP10)**



GenCore version 5.1.1.3  
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM nucleic - protein search, using frame\_plus\_n2p model  
Run on: December 2, 2002, 18:23:15 ; Search time 55.5 Seconds  
(without alignments)  
5282.007 Million cell updates/sec

Title: US-09-896-791B-2  
Perfect score: 2010  
Sequence: 1 gaattcgacagggccat.....aaaaaaaaaacatcgccgc 1100

Scoring table: BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 1816940

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters: -DEV=rlp  
-MODE=frame+n2p.model -DEV=rlp  
-Q/cgn2\_1/USPTO.spool/US09896791/runat\_02122002\_152657\_5564/app\_query.fasta\_1.1287  
-DB=A\_Geneseq\_101002 -QFMT=fastan -SUFFIX=rag -MINMATCH=0.1 -LOOPCL=0  
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi  
-LIST=45 -DOCALIGN=200 -THR\_SCORE=pct -THR\_MAX=100 -THR\_MIN=0 -ALIGN=15  
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US09896791 @CN\_1.1.42 @runat\_02122002\_152657\_5564 -NCPU=6 -ICPU=3  
-NO\_XLPXY -NO\_MMAP -LARGEQUERY -NEG\_SCORES=0 -WAIT -LONGLOG -DEV\_TIMEOUT=120  
-WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOPOP=6 -FGAPEXT=7  
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : A\_Geneseq\_101002.\*  
1: /SID2/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.\*  
2: /SID2/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.\*  
3: /SID2/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.\*  
4: /SID2/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.\*  
5: /SID2/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.\*  
6: /SID2/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.\*  
7: /SID2/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.\*  
8: /SID2/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.\*  
9: /SID2/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.\*  
10: /SID2/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.\*  
11: /SID2/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.\*  
12: /SID2/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.\*  
13: /SID2/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.\*  
14: /SID2/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.\*  
15: /SID2/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.\*  
16: /SID2/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.\*  
17: /SID2/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.\*  
18: /SID2/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.\*  
19: /SID2/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.\*  
20: /SID2/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.\*  
21: /SID2/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.\*  
22: /SID2/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.\*  
23: /SID2/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	1636	81.4	307	23	AAU75902	Mouse inhibitory P
2	1045	52.0	662	20	AAV06295	Mouse transcriptio
3	1011.5	50.3	632	22	AA93326	Human protein sequ
4	982.5	48.9	667	23	AAE24222	Human HIF-3 protei
5	978	48.7	407	23	ABG66737	Human novel polype
6	753.5	37.5	648	22	AA94934	Human protein sequ
7	672	33.4	199	22	AA93710	Human protein sequ
8	610.5	30.4	810	23	ABB57270	Mouse ischaemic co
9	606.5	30.2	245	21	AAV94627	HIF-lalpa variant
10	606.5	30.2	330	21	AAV94628	HIF-lalpa variant
11	606.5	30.2	642	21	AAV84168	A variant of human
12	606.5	30.2	652	21	AAV94629	HIF-lalpa variant
13	606.5	30.2	669	21	AAV84167	A variant of human
14	606.5	30.2	697	21	AAV84166	A variant of human
15	606.5	30.2	701	21	AAV84173	A variant of human
16	606.5	30.2	710	21	AAV84172	A variant of human
17	606.5	30.2	724	21	AAV84171	A variant of human
18	606.5	30.2	749	21	AAV84170	A variant of human
19	606.5	30.2	789	21	AAV84169	A variant of human
20	606.5	30.2	813	21	AAV94636	HIF-lalpa variant
21	606.5	30.2	826	18	AAW06557	Human hypoxia indu
22	606.5	30.2	826	19	AAW80418	Amino acid sequenc
23	606.5	30.2	826	20	AAV06289	Human transcriptio
24	606.5	30.2	826	21	AAV94640	Human hypoxia-indu
25	606.5	30.2	826	21	AAV69407	A wild type human
26	606.5	30.2	826	22	AAV76854	Human lung tumour
27	606.5	30.2	826	23	AAE24212	Human HIF-lalpa p
28	606.5	30.2	826	23	AAV77602	Human hypoxia-indu
29	606.5	30.2	826	23	AAV77607	Human hypoxia-indu
30	606.5	30.2	826	23	AAV77608	Human hypoxia-indu
31	606.5	30.2	826	23	AAV77609	Human hypoxia-indu
32	606.5	30.2	826	23	AAV77610	Human hypoxia-indu
33	606.5	30.2	826	23	AAV77611	Human hypoxia-indu
34	606.5	30.2	826	23	AAV77612	Human hypoxia-indu
35	606.5	30.2	826	23	AAV77613	Human hypoxia-indu
36	606.5	30.2	826	23	AAV77615	Human hypoxia-indu
37	606.5	30.2	826	23	AAV77616	Human hypoxia-indu
38	606.5	30.2	826	23	AAV77617	Human hypoxia-indu
39	606.5	30.2	826	23	AAV77618	Human hypoxia-indu
40	606.5	30.2	826	23	AAV77619	Human hypoxia-indu
41	606.5	30.2	826	23	AAV77620	Human hypoxia-indu
42	606.5	30.2	826	23	AAU85509	Clone #19080 of lu
43	606.5	30.2	826	23	AAU69409	Lung small cell ca
44	606.5	30.2	826	23	AAU69409	Murine endothelial
45	583	29.0	875	19	AAW37098	

ALIGNMENTS

RESULT 1  
AAU75902  
ID AAU75902 standard; Protein; 307 AA.  
XX AAU75902;  
XX  
XX  
DT 08-MAY-2002 (first entry)  
XX  
DE Mouse inhibitory PAS domain (IPAS) protein.  
XX  
KW Inhibitory PAS domain protein; IPAS; cardiant; vasotropic;  
KW cardiovascular; cerebroprotective; ophthalmological; HIF-lalpa;  
KW hypoxia-induced factor lalpa; vascular endothelial growth factor;  
KW angiogenesis; ischaemic cardiovascular lesion; stroke;  
KW diabetic microvascular disease; tumour; mouse.  
XX  
OS Mus musculus.  
XX  
XX  
PN WO200202609-A1.

10-JAN-2002.  
 19-JUN-2001; 2001WO-SE01387.  
 06-JUL-2000; 2000SE-0002551.  
 (BIOV-) BIOVITRUM AB.  
 Berkenstam A, Bertilsson G, Poellinger L;  
 WPI; 2002-164523/21.  
 N-PSDB; ABK14502.  
 New nucleic acid encoding inhibitory PAS domain protein, useful for  
 identifying specific inhibitors for treating e.g. angiogenesis or  
 tumour growth  
 Claim 3; Fig 1; 44pp; English.  
 The invention describes an isolated nucleic acid encoding the  
 biologically active inhibitory PAS domain protein or its functionally  
 equivalent modifications. IPAS forms a non-functional heterodimeric  
 complex with HIF-1alpha (hypoxia-induced factor 1alpha), impairing  
 interaction between HIF-1alpha and hypoxia-response elements in genes,  
 e.g. the gene for vascular endothelial growth factor, so contributes to  
 control of hypoxic signalling. The nucleic acid and its encoded  
 polypeptides, are used to identify agents that activate expression of  
 the gene or stimulate activity of the protein. These agents are useful  
 for inhibiting angiogenesis, particularly where associated with ischaemic  
 cardiovascular lesions, stroke or diabetic microvascular diseases, and  
 tumour growth. This is the amino acid sequence of the mouse inhibitory  
 PAS domain protein (IPAS), described in the method of the invention.  
 Sequence 307 AA;

Alignment Scores:  
 Pred. No.: 9,76e-139 Length: 307  
 Score: 1636.00 Matches: 307  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 81.39% Indels: 0  
 DB: 23 Gaps: 0

US-09-896-791B-2 (1-1100) x AAU75902 (1-307)

QY 19 ATGGCGTTGGCGTGCAGCGCTGAGTGCAGACACCGAGTGGCGAAGGAGAGTCCCGG 78  
 Db 1 MetAlaLeuGlyLeuGlnArgValArgSerAsnThrGluLeuArgLysGluLysSerArg 20  
 QY 79 GACGCGCGCGCAGCGCGCGCAGGAGGAGGAGTGTGTACACAGTGGCGCACACT 138  
 Db 21 AsplaalaargSerArgSerGlnGluThrGluValLeuYrGlnLeuAlaHisThr 40  
 QY 139 CTGCGCTTTGGCGCGCGCGTACGCGCACCTGGACAGGCGCTCCATCATCGCGCTCACA 198  
 Db 41 LeuProPheAlaArgGlyValSerAlaHisLeuAspLysAlaSerIleMetArgLeuThr 60  
 QY 199 ATCAGCTACCTGCAGTCCAGCGCGCTGCGCAGCAGTGGGAAAGGGGAGAGCCACT 258  
 Db 61 IleSerTyrLeuArgMetHisArgLeuLysAlaAlaGlyLysArgGlyArgAlaThr 80  
 QY 259 GGAGCGCTGTACTGAGGCGCGCGTGGAGGTTTCGTATGCTACCTACCGCGGAGGAGA 318  
 Db 81 GlyArgLeuLeuProGluGlyProGlyGlyPheArgHisGlyThrHisArgGlyArg 100  
 QY 319 CATGGCTTACCTGTGCGAAATGTACAGACACCTGGCGCTCAGTCAGTGGACCTCTGT 378  
 Db 101 HisGlyLeuProValGlyLysCysGlnGlnAlaProGlyProGlnSerValAspLeuCys 120  
 QY 379 TCCTCTCCCTGATACATACCCCACTCTGGTACCAATTTCTCTCGGAGCTCATTTGGA 438  
 Db 121 SerSerSerLeuIleHisAsnProThrProGlyThrAsnPheSerLeuGluLeuIleGly 140

QY 439 CACAGTATCTTTGATTTTATCCATCCCTGTGACCAAGAGGAACTTCAAGAGCCCTGACC 498  
 Db 141 HisSerIlePheAspPheIleHisProCysAspGlnGluLeuGlnAspAlaLeuThr 160  
 QY 499 CCAGAGCGGAACTGTCAAGAAGAACTGGAAGCCCAACAGAGCGCCACTTTTCCCTG 558  
 Db 161 ProArgProAsnLeuSerLysLysLeuGluAlaProThrGluArgHisPheSerLeu 180  
 QY 559 CGAATGAAGACGCTCACAGCAGAGGCGCGCAGCTCAACCTCAAGAGCGCCACCTGG 618  
 Db 181 ArgMetLysSerThrLeuThrSerArgGlyArgThrLeuAsnLeuLysAlaAlaThrTrp 200  
 QY 619 AAGGTGCTGCACTGCTCAGGACATATAGGGCCCTACAGGCCCTTGACAGACTTCCCT 678  
 Db 201 LysValLeuHisCysSerGlyHisMetArgAlaThrLysProProAlaGlnThrSerPro 220  
 QY 679 GCCGGAGCGCTCGCTCCGAGCCCTCCCTGCAATGCTGGTGTATCTGTGAAGCCATC 738  
 Db 221 AlaGlySerProArgSerGluProProLeuGlnCysLeuValLeuIleCysGluAlaIle 240  
 QY 739 CCCAGCTCCCTTCCAGATGGTGTACTCTGGGTCTTCCACAGGAGAACTCCCATC 798  
 Db 241 ProGlnLeuProPheHisAspGlyAlaThrLeuGlyLeuProGlnGluLysThrProIle 260  
 QY 799 TCTACCTTATTACCCCTCTTTGGAAGGCACTACTTTGTCTTGTCAAGAGGTGGCTGTT 858  
 Db 261 SerThrLeuPheThrProLeuTrpLysAlaLeuLeuCysLeuValLysArgTrpProVal 280  
 QY 859 CAGGTGCTACAGGGGAAAGGACTGAATCTCTCTCCCTCATGGGTGTGTGGGCCCTT 918  
 Db 281 GlnValLeuGlnGlyLysGlyThrGluSerSerLeuProSerTrpValLeuTrpAlaLeu 300  
 QY 919 AACCGGAAAATGTCTCTGCG 939  
 Db 301 AsnArgLysAsnCysProGly 307

RESULT 2  
 AAY06295  
 ID AAY06295 standard; Protein; 562 AA.  
 XX AC AAY06295;  
 XX XX  
 DT 23-AUG-1999 (first entry)  
 DE Mouse transcription regulator MOP7.  
 XX KW MOP7; member of the PAS superfamily; bHLH-PAS; mouse;  
 KW transcriptions regulator; hypoxia inducible factor 3 alpha.  
 XX OS Mus musculus.  
 XX PN WO928464-A2.  
 XX PD 10-JUN-1999.  
 XX PF 27-NOV-1998; 98WO-US25314.  
 XX PR 28-NOV-1997; 97US-0066863.  
 XX PA (WISC ) WISCONSIN ALUMNI RES FOUND.  
 XX PI Bradford CA, Gu YZ, Hogenesch JB;  
 XX WPI; 1999-371120/31.  
 DR N-PSDB; AAX58986.  
 XX PT Developmental signal transduction associated proteins  
 XX PS Claim 6; Page 101; 106pp; English.  
 CC The present sequence represents mouse MOP7, a novel member of the  
 CC PAS superfamily, where PAS stands for PER/ARNT/SIM domains. MOP7

CC cDNA (see AAX38986) was identified in a search of murine ESTs designed  
 CC to identify bHLH-PAS proteins, and by RACE amplification of lung  
 CC cDNA. MOP7 was characterised as hypoxia inducible factor 3 alpha  
 CC (HIF 3 alpha). Its expression profile is distinct from that of  
 CC HIF 1 alpha (see AAY06289), HIF 2 alpha (see AAY06290), MOP3 (see  
 CC AAY06291), Ah receptor and Ah receptor nuclear translocator (ARNT),  
 CC suggesting a different functional role. MOP7 probably regulates  
 CC the same genes as HIF 1 alpha and 2 alpha, as evidenced by its  
 CC dimerisation with the same partners (ARNT, MOP3) and recognition  
 CC of the same core response element. MOP7 may have a functional  
 CC role associated with response to low oxygen in the tissues in  
 CC which it is expressed. The invention provides novel MOPs 2-9  
 CC nucleic acids (see AAX58981-88) and proteins (see AAY06289-97).  
 CC These are useful in a variety of research, diagnostic and  
 CC therapeutic applications. Several of the MOPs are alpha-class  
 CC hypoxia-inducible factors. Others are involved in circadian signal  
 CC transduction.

XX Sequence 662 AA;

Alignment Scores:  
 Pred. No.: 2,03e-85 Length: 662  
 Score: 1045.00 Matches: 218  
 Percent Similarity: 85.88% Conservative: 1  
 Best Local Similarity: 85.49% Mismatches: 5  
 Query Match: 51.99% Indels: 31  
 DB: 20 Gaps: 3

US-09-896-791B-2 (1-1100) x AAY06295 (1-662)

QY 43 AGGTGACACCGAGTGGCGAAGAGAGTCCGGGACGCGCGCGCGCGCGCAGC 102  
 Db 7 ArgSerAsnThrGluLeuArgLysGluLysSerArgAspAlaAlaArgSerArgSer 26  
 QY 103 CAGGACAGGAGTGTGTACACAGCTGGCGCACACTCTGCCCTTTCGGCGCGCGTACG 162  
 Db 27 GlnGluThrGluValLeuTyrGlnLeuAlaHisThrLeuProPheAlaArgGlyValSer 46  
 QY 163 GCGACCTGGACAGCGCTCATCATCGCCCTCACATACAGCTACCTGGCGATGCACCGC 222  
 Db 47 AlaHisLeuAspLysAlaSerIleMetArgLeuThrIleSerTyrLeuArgMetHisArg 66  
 QY 223 CTCCTGGCAGCAGGTGGAA-----AAAGGGGAGAGCCACTGGACGCCCTG 267  
 Db 67 LeuCysAlaAlaGly-GluTrpAsnGlnValGluLysGlyGlyGluProLeuAspAlaCy 86  
 QY 268 CTACCTGAGGCCCTGGAGGTTTCCTCATGTACTACCGCGGAGGAGACATGGCTTA 327  
 Db 86 sTyrLeuLysAlaLeuGluGlyPheValMetValLeuThrAlaGluGlyAspMetAlaTy 106  
 QY 328 CTTGTCGGAATGTCAGAACGACCTGGGCGCTCAGTCACTGGACCTCTCTCTCTCTCC 387  
 Db 106 rLeuSerGluAsnValSerLysHisLeuGlyLeuSerGln----- 119  
 QY 388 CTGATACATAACCCCACTCTCTGTGTACCAATTTCTCTCTGTGAGCTCATTTGGACAGTATC 447  
 Db 120 -----Leu--GluLeuIleGlyHisSerIle 127  
 QY 448 TTTGATTTATTCATCCCTGTGACCAAGAGGAATCTTCAAGACGCCCTGACCCCGCGCG 507  
 Db 128 PheAspPheIleHisProCysAspGlnGluLeuGlnAspAlaLeuThrProArgPro 147  
 QY 508 AACCTGTCAAAGAAGCTGGAAGCCCAACAGAGCGCCACTTTTCCCTGCGCAATGAAG 567  
 Db 148 AsnLeuSerLysLysLeuGluAlaProThrGluArgHisPheSerLeuArgMetLys 167  
 QY 568 AGCAGCTTACACGAGAGGCGCGCAGCTCAACCTCAAGCGCGCCACCTGGAAAGTGGCTG 627  
 Db 168 SerThrLeuThrSerArgGlyArgThrLeuAsnLeuLysAlaAlaThrTrpLysValLeu 187  
 QY 628 CACTGCTCAGGACATATGAGGGGCTTCAAGCCCGCTGCACAGACTTCCCTGCGCGGAGC 687  
 Db 188 HisCysSerGlyHisMetArgAlaTyLysProAlaGlnThrSerProAlaGlySer 207

QY 688 CTTGCTCGAGCCTCCCTGCAATGCTGCTTATCTGTGAAGCCATCCCC----- 741  
 Db 208 ProArgSerGluProProLeuGlnCysLeuValLeuIleCysGluAlaIleProHisPro 227  
 QY 742 -----CAGTCTCCCTTCCACGATGGTGTACTCTG 771  
 Db 228 AlaSerLeuGluProProLeuGlyArgGlyAlaPheLeu 240

RESULT 3

AAB93326

ID AAB93326 standard; Protein; 632 AA.

AC AAB93326;

DT 26-JUN-2001 (first entry)

DE Human protein sequence SEQ ID NO:12422.

XX Human; primer; detection; diagnosis; antisense therapy; gene therapy.

XX Homo sapiens.

XX EP1074617-A2.

XX 07-FEB-2001.

XX 28-JUL-2000; 2000EP-0116126.

XX 29-JUL-1999; 99JP-0248036.

XX 27-AUG-1999; 99JP-0300253.

XX 11-JAN-2000; 2000JP-0118776.

XX 02-MAY-2000; 2000JP-0183767.

XX 09-JUN-2000; 2000JP-0241899.

XX (HELI-) HELIX RES INST.

XX Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;

PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;

XX WPI; 2001-318749/34.

DR

XX

PT

PT

PT

PT

XX

PS

XX

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

XX

SQ

Sequence 632 AA;







QY 592 ACCTCAACCTCAAGCGCCACCTGGAAGGTGCTGCACTGCTCAGGACATATGAGGGCC 651  
DB 122 ThrLeuAsnLeuLysAlaAlaThrTrpLysValLeuAsnCysSerGlyHisMetArgAla 141  
QY 652 TACAAGCCCCCTGCACAGACTTCCCTGCGGGAGCCCTCGCTCCAGGCTCCCTGCA 711  
DB 142 TyrLysProAlaGlnThrSerProAlaGlySerProAspSerGluProProLeuGln 161  
QY 712 TGCTGTGCTGCTATCTGTCAGCCATCCCCAGCTCCCTTCCACGATGCTGCT 765  
DB 162 CysLeuValleuIleCysGluAlaIlePro-----HisProGlySer 175  
RESULT 7  
AAB93710  
ID AAB93710 standard; Protein; 199 AA.  
AC AAB93710;  
DT 26-JUN-2001 (first entry)  
XX Human protein sequence SEQ ID NO:13303.  
DE Human; primer; detection; diagnosis; antisense therapy; gene therapy.  
KW Homo sapiens.  
OS EP1074617-A2.  
PN 07-FEB-2001.  
PD 28-JUL-2000; 2000EP-0116126.  
PF 29-JUL-1999; 99JP-0248036.  
PR 27-AUG-1999; 99JP-0300253.  
PR 11-JAN-2000; 2000JP-0118776.  
PR 02-MAY-2000; 2000JP-0183767.  
PR 09-JUN-2000; 2000JP-0241899.  
XX (HELI-) HELIX RES INST.  
XX Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;  
PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;  
XX WPI; 2001-318749/34.  
XX Primer sets for synthesizing polynucleotides, particularly the 5602  
PT full-length cDNAs defined in the specification, and for the detection  
PT and/or diagnosis of the abnormality of the proteins encoded by the  
PT full-length cDNAs -  
XX Claim 8; SEQ ID 13303; 2537pp + CD ROM; English.  
PS The present invention describes primer sets for synthesizing 5602  
XX full-length cDNAs defined in the specification. Where a primer set  
CC comprises: (a) an oligo-dr primer and an oligonucleotide complementary  
CC to the complementary strand of a polynucleotide which comprises one of  
CC the 5602 nucleotide sequences defined in the specification, where the  
CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination  
CC of an oligonucleotide comprising a sequence complementary to the  
CC complementary strand of a polynucleotide which comprises a 5'-end  
CC sequence and an oligonucleotide comprising a sequence complementary to a  
CC polynucleotide which comprises a 3'-end sequence, where the  
CC oligonucleotide comprises at least 15 nucleotides and the combination of  
CC the 5'-end sequence/3'-end sequence is selected from those defined in  
CC the specification. The primer sets can be used in antisense therapy and  
CC in gene therapy. The primers are useful for synthesizing polynucleotides,  
CC particularly full-length cDNAs. The primers are also useful for the  
CC detection and/or diagnosis of the abnormality of the proteins encoded by  
CC the full-length cDNAs. The primers allow obtaining of the full-length  
CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and  
CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to  
CC AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632

CC represent oligonucleotides, all of which are used in the exemplification  
XX of the present invention.  
SQ Sequence 199 AA;  
Alignment Scores:  
Pred. No.: 5.43e-52 Length: 199  
Score: 672.00 Matches: 141  
Percent Similarity: 64.19% Conservative: 6  
Best Local Similarity: 61.57% Mismatches: 14  
Query Match: 33.43% Indels: 68  
DB: 22 Gaps: 2  
US-09-896-791B-2 (1-1100) x AAB93710 (1-199)  
QY 83 CGGCGCGAGCGCGCGCAGGACGAGGAGTGTACCAAGCTGGCGCACACTCTGC 142  
DB 2 ArgProAlaAlaGlyAlaAlaArgArgProArgCysCysThrSerTrpLeuThrArgCys 21  
QY 143 CCTTTGCGGCGGCGTCAAGCGCACCTGCGAGGAGGCTCCATCATCGCCCTCACAATCA 202  
DB 22 ProSerProAlaAlaSerAlaProThrTrpThrArgProLeuSerCysAlaSerProSer 41  
QY 203 GCTACCTGCGCATGACCGCCTCTGCGCAGCAGGCTGGAAGGGGAGAGCCACTGGAC 262  
DB 42 AlaThrCysAlaCysThrAlaSerAlaProGln----- 52  
QY 263 GCCTGCTACCTGAAGGCCCTGGAGGCTTTCGTCATCGTACTACCCGCCGAGGAGACATG 322  
DB 52 ----- 52  
QY 323 GCTTACCTGTCGAAAAATGTCAGCAAGCACCTGGGCTCAGTCAGTGGACCTCTGTTCT 382  
DB 52 ----- 52  
QY 383 CCTCCTGATACATAACCCACTCTCTGGTACCAATTTCTCTGAGCTCATFTGGACACA 442  
DB 53 -----Leu-GluLeuIleGlyHis 59  
QY 443 GTATCTTTGATTTTATCCATCCCTGTGACCAAGAGGAAGTTCAGAGCGCTGACCCCA 502  
DB 59 efilePheAspPheIleHisProCysAspGlnGluLeuGlnAspAlaLeuThrProG 79  
QY 503 GCGCGAACCTGTCAAGAAAGAGCTGGAAGCCCAAGAGCGCCACTTTTCCCTCGCAA 562  
DB 79 InGlnThrLeuSerArgArgLysValGluAlaProThrGluArgCysPheSerLeuArgM 99  
QY 563 TGAAGAGCAGCTTACACGAGGCGCAGCGCTCAACCTCAAGCGCGCACCTGGAAGG 622  
DB 99 etLysSerThrLeuThrSerArgGlyArgThrLeuAsnLeuLysAlaAlaThrTrpLys 119  
QY 623 TGCTGCATGCTCAGGACATATAGGGGCTACAGGCCCTTACAGCGCTGACACTTCCCTGCCG 682  
DB 119 alLeuAsnCysSerGlyHisMetArgAlaTrpLysProAlaGlnThrSerProAla 139  
QY 683 GGAGCCCTCGCTCCGAGCCTCCCTGCAATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 742  
DB 139 lySerProAspSerGluProProLeuGlnCysLeuValleuIleCysGluAlaIlePro- 158  
QY 743 AGCTCCCTTCCACGATGCTGCT 765  
DB 159 -----HisProGlySer 162  
RESULT 8  
AAB57270  
ID ABB57270 standard; Protein; 810 AA.  
XX AAB57270;  
XX 07-MAR-2002 (first entry)  
XX Mouse ischaemic condition related protein sequence SEQ ID NO:760.  
DE  
XX





CC The invention relates to isolated variants of HIF-1alpha, such as that  
CC represented by the present sequence. The variants are useful for  
CC identifying compounds capable of modulating the function of a functional  
CC domain of human HIF-1alpha. The method comprises contacting a candidate  
CC compound with a cell expressing a HIF-1alpha variant conjugated to a  
CC molecular probe. The localization of the probe can be detected in the  
CC cell. The Aequorea victoria green fluorescent protein can be used as the  
CC molecular probe. The compounds are useful for the regulation of  
CC HIF-1alpha target genes, such as those involved in the regulation of  
CC angiogenesis, erythropoiesis and glycolysis.

XX Sequence 245 AA;

Alignment Scores:  
Pred. No.: 4 57e-46 Length: 245  
Score: 606.50 Matches: 129  
Percent Similarity: 68.07% Conservativeness: 33  
Best Local Similarity: 54.20% Mismatches: 48  
Query Match: 30.17% Indels: 28  
DB: 21 Gaps: 3

US-09-896-791B-2 (1-1100) x AAY94627 (1-245)

QY 49 AACACCGAGCTCGGAGGAGAGTTCGGGGGAGCGGGCCGCGCGGCGCAGCAGGAG 108  
Db 14 SerSerGluArgArgLysGluLysSerArgAspAlaAlaArgSerArgSerLysGlu 33  
QY 109 ACGGAGGTCTTACACGTGGCGCACACTCTGCCCTTTGGCGGCGCTCAGCGGCAC 168  
Db 34 SerGluValPheTyrGluLeuAlaHisGlnLeuProLeuProHisAsnValSerHis 53  
QY 169 CTGGACAGGCTCCATCATGCGCTCAACATCAGTCTGCTGCGATGCGCGCTCTGC 228  
Db 54 LeuAspLysAlaSerValMetArgLeuThrIleSerTyrLeuArgValArgLysLeu 73  
QY 229 GCAGCAGGTGGAAGAGGGGAGAG-----CCACTGGAGCGCTGTACTAC 273  
Db 74 AspAlaGly-AspLeuAspIleGluAspAspMetLysAlaGlnMetAsnCysPheTyrLe 93  
QY 274 GAAGCGCTGGAGGTTCGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 333  
Db 93 uLysAlaLeuAspGlyPheValMetValLeuThrAspAspGlyAspMetIleTyrIleSe 113  
QY 334 GGAATATGTCAGACAGCAGCTGGGCTCAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 393  
Db 113 rAspAsnValAsnLysTyrMetGlyLeuThrGln----- 124  
QY 394 CATACCCCATCTCTGGTACCAATTTCTCTGAGCTCATTTGGACAGTATCTTTGAT 453  
Db 125 -----Phe--GluLeuThrGlyHisSerValPheAsp 134  
QY 454 TTTATCATCTCTGACCAAGAGGAGTTCAGAGCGCTGACCGCGCGGCGGCGGCGGCGG 513  
Db 135 PheThrHisProCysAspHisGluGluMetArgGlnMetLeuThrHisArgAsnGlyLeu 154  
514 TCAAGAGAAGCTGAAGCCCAACAGAGCGCGCTTTTCCCTGCGAATGAAGACGAG 573  
Db 155 ValLysLysGlyGluGlnAsnThrGlnArgSerPhePheLeuArgMetLysCysThr 174  
QY 574 CTCACAGCAGAGCGCGCGCTCAACCTCAAGCGGCGGCGGCGGCGGCGGCGGCGGCGG 633  
Db 175 LeuThrSerArgGlyArgThrMetAsnIleLysSerAlaThrTrpLysValLeuHisCys 194  
QY 634 TCAGGACATATAGGCGCTTACAGCGCGCTGCGACAGACTTCCCTGCGGCGGCGGCGGCGG 693  
Db 195 ThrGlyHisIleHisValTyrAspThrAsnSerAsnGlnProGlnCysGly-----Tyr 212  
QY 694 TCCGAGCTCCCTGCAATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 741  
Db 213 LysLysProProMetThrCysLeuValLeuIleCysGluProIlePro 228  
RESULT 10  
AAY94628

ID AAY94628 standard; protein; 330 AA.  
XX  
AC AAY94628;  
XX  
DT 15-AUG-2000 (first entry)  
XX  
- DE HIF-1alpha variant protein sequence HIF-1alpha/1-330.  
XX  
KW Hypoxia-inducible factor 1alpha; HIF-1alpha; PAS-B; N-TAD; C-TAD;  
KW regulation; angiogenesis; erythropoiesis; glycolysis; human.  
XX  
OS Homo sapiens.  
XX  
PN WO200029437-A1.  
XX  
PD 25-MAY-2000.  
XX  
PF 11-NOV-1999; 99WO-SE02053.  
XX  
PR 13-NOV-1998; 98SE-0003891.  
XX  
PA (PHAA ) PHARMACIA & UPJOHN AB.  
XX  
PI Berkenstam A, Poellinger L;  
XX  
DR WPI; 2000-399715/34.  
XX  
PT Human hypoxia-inducible factor 1alpha variants for identifying compounds  
PT that modulate its functional domain and regulate genes involved in  
PT angiogenesis, erythropoiesis -  
XX  
PS Claim 15; Page 69-70; 87pp; English.

CC This sequence represents a fragment of the hypoxia-inducible factor  
CC (HIF)-1alpha amino acid sequence. The mechanism of action of HIF-1alpha  
CC is a multi-step process which includes hypoxia-dependent nuclear import  
CC and activation of the transactivation domain. The HIF-1alpha consists of  
CC a number of functional domains including a PAS-B (Per, Arnt, Sim) domain  
CC located in human HIF-1alpha between amino acids 173 and 390, a C-terminal  
CC nuclear localization sequence located at amino acids 718-584, a  
CC transactivator domain (N-TAD) located between amino acids 531 and 584,  
CC and a second transactivator domain (C-TAD) located between 813 and 826.  
CC The invention relates to isolated variants of HIF-1alpha, such as that  
CC represented by the present sequence. The variants are useful for  
CC identifying compounds capable of modulating the function of a functional  
CC domain of human HIF-1alpha. The method comprises contacting a candidate  
CC compound with a cell expressing a HIF-1alpha variant conjugated to a  
CC molecular probe. The localization of the probe can be detected in the  
CC cell. The Aequorea victoria green fluorescent protein can be used as the  
CC molecular probe. The compounds are useful for the regulation of  
CC HIF-1alpha target genes, such as those involved in the regulation of  
CC angiogenesis, erythropoiesis and glycolysis.

XX SQ Sequence 330 AA;

Alignment Scores:  
Pred. No.: 5e-46 Length: 330  
Score: 606.50 Matches: 129  
Percent Similarity: 68.07% Conservativeness: 33  
Best Local Similarity: 54.20% Mismatches: 48  
Query Match: 30.17% Indels: 28  
DB: 21 Gaps: 3

US-09-896-791B-2 (1-1100) x AAY94628 (1-330)

QY 49 AACACCGAGCTCGGAGGAGAGTTCGGGGGAGCGGGCCGCGCGGCGCAGCAGGAG 108  
Db 14 SerSerGluArgArgLysGluLysSerArgAspAlaAlaArgSerArgSerLysGlu 33  
QY 109 ACGGAGGTCTGTACACGTGGCGCACACTCTGCCCTTTGGCGGCGGCGGCGGCGGCG 168  
Db 34 SerGluValPheTyrGluLeuAlaHisGlnLeuProLeuProHisAsnValSerHis 53







CC for increasing expression of a hypoxia inducible gene in a cell. They  
CC is also useful for providing constitutive expression of a hypoxia  
CC inducible factor in a cell, and for reducing or preventing hypoxia or  
CC ischemia related damage. The variant HIF-1alpha polypeptides are useful  
CC for providing prophylactic therapy for inducing the level of  
CC angiogenesis in tissues of patients at risk of coronary artery disease  
CC or ischemic tissue damage.  
CC note: this sequence does not appear in the specification; it was created  
CC using information provided.  
XX Sequence 697 AA;

Alignment Scores:  
Pred. No.: 6,28e-46 Length: 697  
Score: 606.50 Matches: 129  
Percent Similarity: 68.07% Conservatives: 33  
Best Local Similarity: 54.20% Mismatches: 48  
Query Match: 30.17% Indels: 28  
DB: 21 Gaps: 3

US-09-896-791B-2 (1-1100) x AAY84166 (1-697)

QY 49 AACACCGAGTCGGGAAGAGAGTAGTCGGGGACGGCGCGCGAGCGCGCCAGCCAGGAG 108  
Db 14 SerSerGluArgArgLysGluLysSerArgAspAlaAlaArgSerArgArgSerLysGlu 33  
QY 109 ACGAGGTGCTGCTACACAGTGGCGGCACACTCTGCCCTTTTCGCGCGCGCTCAGCGGCAC 168  
Db 34 SerGluValPheTyrGluLeuAlaHisGlnLeuProHisAsnValSerSerHis 53  
QY 169 CTGGACAAGGCTCCATCATGCGCCCTCACAACTACGTACTGCGCATGCGCCCTTCGC 228  
Db 54 LeuAspLysAlaSerValMetArgLeuThrIleSerTyrLeuArgValArgLysLeuLeu 73  
QY 229 GCAGAGGTGGAAAAAGGGGAGAG-----CCACTGGAGCGCTGCTACCT 273  
Db 74 AspAlaGly-AspLeuAspIleGluAspMetLysAlaGlnMetAsnCysPheTyrLe 93  
QY 274 GAAGGCCCTGGAGGTTTCTGCTACTGCTACTCACCAGCGGGGAGACATGCTTACCTGTC 333  
Db 93 uLysAlaLeuAspGlyPheValMetValLeuThrAspAspGlyAspMetIleTyrIleSe 113  
QY 334 GAAAAATGTGACGAAGCACTGGGCGCTCAGTCAGTGAGCTCTGTCTCTCCCTCCCTGATA 393  
Db 113 rAspAsnValAsnLysTyrMetGlyLeuThrGln----- 124  
QY 394 CATACCCCACTCCTGGTACCAATTTCTCTCTGAGAGCTCATGTGGACACAGTATCTTTGAT 453  
Db 125 -----Phe--GluLeuThrGlyHisSerValPheAsp 134  
QY 454 TTTATCATCCCTGTGACCAAGAGAACTTCAAGAGCGCTGACCCCGAGCGGACCTG 513  
Db 135 PheThrHisProCysAspHisGluGluMetArgGluMetLeuThrHisArgAsnGlyLeu 154  
QY 514 TCAAGAAGAAGCTGGAAGCCCAACAGAGCGCCACTTTTCCCTGCGAATGAAGAGCACG 573  
Db 155 ValLysLysGlyLysGluGlnAsnThrGlnArgSerPhePheLeuArgMetLysCysThr 174  
QY 574 CTCACAGCAGAGGGCGCGCTCAACCTCAACGAGCGCCACCTGGAGGTGCTGCACTGC 633  
Db 175 LeuThrSerArgGlyArgThrMetAsnIleLysSerAlaThrTrpLysValLeuHisCys 194  
QY 634 TCAGGACATATGAGGCGCTTACAGCCCGCTGCACACACTTCCCTCGCGGAGCCCTCGC 693  
Db 195 ThrGlyHisIleHisValTyrAspThrAsnSerAsnGlnProGlnCysGly-----Tyr 212  
QY 694 TCCGAGCTCCCTCGCAATCGCTGGTGTATCTGTGAAGCCATCC 741  
Db 213 LysLysProPromethrCysLeuValLeuIleCysGluProIlePro 228

RESULT 15  
AAY84173  
ID AAY84173 standard; Protein; 701 AA.

XX AAY84173;  
XX 03-JUL-2000 (first entry)  
XX A variant of human hypoxia inducible factor-1 alpha protein.  
XX Human; hypoxia-inducible factor 1 alpha; HIF-1alpha; variant;  
KW hypoxia inducible gene; hypoxia inducible factor; hypoxia;  
KW ischemia related damage; angiogenesis; coronary artery disease;  
KW ischemic tissue damage.  
XX Synthetic.  
OS Homo sapiens.  
XX Key Location/Qualifiers  
FT Misc-difference 426 /note= "this residue is not Ser, and is  
FT preferably Gly"  
FT Misc-difference 427 /note= "this residue is not Thr, and is  
FT preferably Ala"

WO200010578-A1.

02-MAR-2000.

25-AUG-1999; 99WO-US19416.

25-AUG-1998; 98US-0148547.

(UYJO ) UNIV JOHNS HOPKINS SCHOOL MEDICINE.

Semenza GL;

WPI; 2000-246493/21.

Variant forms of hypoxia-inducible factor (HIF)-1 alpha, useful for  
treating hypoxia or ischemia-related tissue damage -

Claim 1; Page -; 96pp; English.

The present sequence represents a variant of hypoxia-inducible factor  
(HIF)-1 alpha, comprising amino acids 1-391 and 517-826 of the wild  
type protein (see AAY69407). The HIF-1alpha variants are stable under  
hypoxic and non-hypoxic conditions. The variants comprises amino acid  
residues 1-391 and 521-826, 549-826, 576-826, 429-826, 469-826, 494-826,  
508-826, 512-826 or 517-826 of the wild type human HIF-1alpha  
polypeptide, in which residues 551 and 552 are not serine and threonine,  
respectively. The HIF-1alpha variant polynucleotide sequences are useful  
for increasing expression of a hypoxia inducible gene in a cell. They  
is also useful for providing constitutive expression of a hypoxia  
inducible factor in a cell, and for reducing or preventing hypoxia or  
ischemia related damage. The variant HIF-1alpha polypeptides are useful  
for providing prophylactic therapy for inducing the level of  
angiogenesis in tissues of patients at risk of coronary artery disease  
or ischemic tissue damage.  
CC note: this sequence does not appear in the specification; it was created  
using information provided.

Sequence 701 AA;

Alignment Scores:  
Pred. No.: 6,29e-46 Length: 701  
Score: 606.50 Matches: 129  
Percent Similarity: 68.07% Conservatives: 33  
Best Local Similarity: 54.20% Mismatches: 48  
Query Match: 30.17% Indels: 28  
DB: 21 Gaps: 3

US-09-896-791B-2 (1-1100) x AAY84173 (1-701)

QY 49 AACACCGAGTCGGGAAGAGAGTAGTCGGGGACGGCGCGCGAGCCGCGCAGGAG 108

```
Db 14 SerSerGluArgArgLysGluLysSerArgAspAlaAlaArgSerArgSerLysGlu 33
QY 109 ACGGAGGTGCTGTACACGCGGCGCACACTGCGCTTGGCGCGCGCTCAGCGCGCAC 168
Db 34 SerGluValPheTyrGluLeuAlaHisGlnLeuProLeuProHisAsnValSerHis 53
QY 169 CTGACAAGGCTCCATCATCGCGCTCACAAATCAGCTACCTGCGCATGCGCGCTCTGC 228
Db 54 LeuAspLysAlaSerValMetArgLeuThrIleSerTyrLeuArgValArgLysLeuLeu 73
QY 229 GCACGAGTGGAAAAGGGGAGAG-----CCACTGGACGCGCTGCTACCT 273
Db 74 AspAlaGly-AspLeuAspIleGluAspAspMetLysAlaGlnMetAsnCysPheTyrLe 93
QY 274 GAAGGCGCTGGAGGTTCGTCTCATGTGTACTCACCGCGCGGAGGAGACATGCTTACCTGTC 333
Db 93 uLysAlaLeuAspGlyPheValMetValLeuThrAspGlyAspMetIleTyrIleSe 113
QY 334 GGAATATGTCAGCAAGCACCTGGGCGCTCAGTCACTGAGTGGACCTCTGCTCCTCCTGATA 393
Db 113 rAspAsnValAsnLysTyrMetGlyLeuThrGln----- 124
QY 394 CATAAACCCCACTCCTGTGTACCAATTCTCTCTGTGAGCTCATTTGGACACAGTATCTTTGAT 453
Db 125 -----Phe--GluLeuThrGlyHisSerValPheAsp 134
QY 454 TTTATCCATCCCTGTGACCAAGAGGAACCTCAAGACGCCCTGACCCCGCGCGAACCTG 513
Db 135 PheThrHisProCysAspHisGluGluMetArgGluMetLeuThrHisArgAsnGlyLeu 154
QY 514 TCAAGAAGAAGCTGGAAGCCCAACAGAGCGCCACTTTTCCCTGCGAATGAAGAGCACG 573
Db 155 ValLysLysGlyLysGluGlnAsnThrGlnArgSerPhePheLeuArgMetLysCysThr 174
QY 574 CTCACAGCAGAGGCGCGCACGCTCAACCTCAAAGCGGCCACCTGGAAGGTGCTGCACTGC 633
Db 175 LeuThrSerArgGlyArgThrMetAsnIleLysSerAlaThrTrpLysValLeuHisCys 194
QY 634 TCAGGACATATGAGGCGCTACAGCCCTTGACAGACTTCCCTCGCGGGAGCCCTCGC 693
Db 195 ThrGlyHisIleHisValTyrAspThrAsnSerAsnGlnProGlnCysGly-----Tyr 212
QY 694 TCCGAGCCTCCCTGCAATGCTGGTGTATCTGTGAGGCCATCCCC 741
Db 213 LysLysProPrometThrCysLeuValLeuIleCysGluProIlePro 228
```

Search completed: December 2, 2002, 20:36:01  
Job time : 63.5 secs